Vision Research 49 (2009) 140–163

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Cortical dynamics of figure-ground segmentation: Shine-through ☆

Gregory Francis

Department of Psychological Sciences, Purdue University, 703 Third Street, West Lafayette, IN 47907-2004, USA

ARTICLE INFO

Article history:
Received 17 March 2008
Received in revised form 20 August 2008

Keywords:
Perceptual organization
Backward masking
Figure-ground segmentation
Occlusion

ABSTRACT

The shine-through effect occurs when a brief offset vernier target is followed by a grating of non-offset vernier elements. Rather than mask the target, this stimulus sequence produces a percept of the target vernier occluding the mask elements. We analyzed the dynamics of the 3D LAMINART model of depth perception and found that it explains the appearance of shine-through for these stimuli. The model explanation proposes that shine-through is due to a combination of false binocular disparity matches between the target and the central element of the mask, and a weakening of between disparity competition due to spatial competition and boundary grouping. Simulations of the model demonstrate that its behavior closely matches empirical data on the properties of shine-through. The model is contrasted with an alternative explanation of shine-through, and novel mask conditions are studied that allow for empirical tests of the model hypotheses.

1. Introduction

Many of the fundamental properties of visual perception involve both space and time. One of the most fundamental questions about *spatial* vision is how the visual system groups together disparate parts of a scene to identify surfaces, objects, and figure-ground distinctions. Although this grouping process has been studied for nearly one hundred years (e.g., Hochberg, 1971; Rock, 1993; Rubin, 1915) there is still no complete description of how it occurs. One of the most fundamental questions about *temporal* vision is how the visual system creates and updates a representation of a changing visual environment. Such properties are often studied with backward masking, where a following stimulus modifies the percept of a leading stimulus. While much is known about backward masking (Breitmeyer & Öğmen, 2006), and related temporal phenomena, there is currently no complete description of how the visual system maintains close temporal synchrony with the visual environment.

In the divide-and-conquer approach of science, these two fundamental aspects of perception have usually been treated separately. Theories of perceptual grouping have mostly focused on static displays, often with variations of the Gestalt laws (e.g., Craft, Schütze, Niebur, & von der Heydt, 2007; Geisler & Super, 2000), but these theories do not consider many temporal aspects of vision. Theories of backward masking (e.g., Breitmeyer, 1984; Francis, 2003; Weisstein, 1972) almost exclusively focus on the rise and fall of neural or perceptual variables and do not consider the perceptual groupings engendered by the spatial layout of stimuli. While many people have recognized that these two aspects of perception need to be brought together (e.g., Enns & Di Lollo, 2000; Francis & Cho, 2008 Weisstein, 1972), it has been difficult to study these phenomena in a systematic way.

Over the past few years, Herzog and colleagues (e.g., Hermens & Herzog, 2007; Herzog, Dependahl, Schmonsees, & Fahle, 2004; Herzog, Fahle, & Koch, 2001; Herzog & Fahle, 2002; Herzog & Koch, 2001) have identified a visual phenomenon that allows them to carefully study the subtle interplay between perceptual grouping and backward masking. The phenomena is called "shine-through" and it demonstrates how grouping changes the perceptual experience of a target and mask temporal sequence.

Fig. 1 schematizes a standard shine-through stimulus sequence. The target stimulus is an offset vernier. The observer's task is usually to judge the offset direction of the vernier. The target is presented very briefly (often around 20 ms). When the target disappears, it is immediately replaced by a 300 ms grating mask that is made of non-offset vernier elements. The temporal sequence of target and mask stimuli is classically defined as backward masking, and with other stimuli the presence of the mask can render the target perceptually invisible.

Despite its short duration, if the target were presented by itself, most observers would find the offset judgment task fairly easy to perform. In contrast to masking studies, with the stimuli in Fig. 1, the target percept is not rendered invisible, but instead the target vernier appears to occlude (shine-through) the grating mask (Herzog & Koch, 2001). Observers sometimes report that the target

☆ Part of this work was completed while the author was a visiting professor at the Laboratory of Psychophysics, Brain Mind Institute, École Polytechnique Fédérale de Lausanne, Switzerland. The author thanks Michael Herzog and Frouke Hermens for useful discussions and comments on the manuscript and Mark Repnow for assistance developing the simulation programming code.

E-mail address: gfrancis@purdue.edu

0042-6989/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.visres.2008.10.002

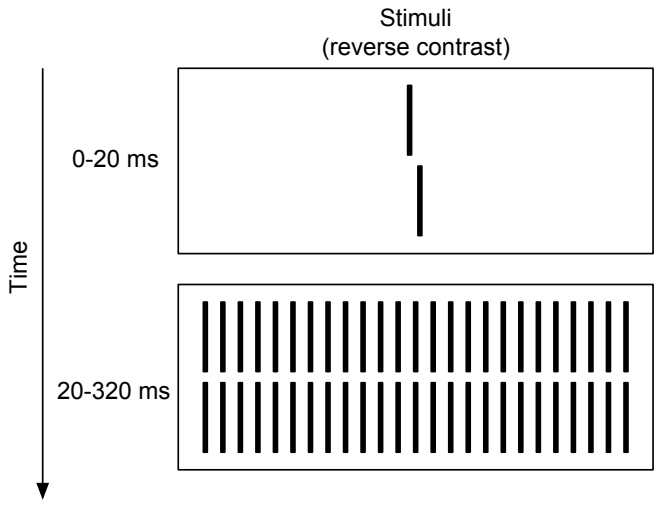

Fig. 1. The stimulus sequence that produces the shine-through effect. A target vernier stimulus is shown for 20 ms, and is then followed by a 300 ms grating mask of non-offset vernier elements. Perceptually, the target vernier appears to occlude the mask elements.

looks brighter, bigger, or closer than the mask elements. It is as if the target is placed in the foreground and occludes the mask elements. When the shine-through effect occurs, the direction of the target vernier offset is easily judged. Thus, the shine-through effect appears to involve both backward masking and perceptual organization (figure-ground distinctions).

In a series of studies described in detail below, Herzog and colleagues showed that perceptual organization of the mask elements played a fundamental role in the creation of the shine-through effect. Herzog and Fahle (2002) concluded that shine-through occurs when the mask elements group together to form an independent and coherent object. Such grouping thereby allows the target to be perceived as a separate entity.

As a verbal description of perceptual grouping, backward masking, and the shine-through effect, we agree with Herzog's characterization of the role of perceptual grouping for the shine-through effect. This verbal description does not, however, characterize how these processes occur in cortical neural circuits. In this paper, we show that a previously created model of visual perception provides a quantitative explanation of the shine-through effect. While conceptually similar to Herzog's description of perceptual grouping, the model hypothesizes a quite different role for perceptual grouping and hypothesizes a number of other stimulus properties that should be necessary for shine-through to occur.

The next section describes the model and explains how it generates a shine-through effect. Following sections then consider empirical properties of shine-through and demonstrate how the model accounts for these data sets. We then contrast the proposed model with a recent alternative explanation of shine-through (Hermens, Herzog, Luksys, Gerstner, & Ernst, 2008). Finally, we describe novel mask stimuli that highlight and test the model's explanation of the shine-through effect. Some of these predictions should allow for a test of the two competing model explanations.

2. Analysis of the 3D LAMINART model

To explain the shine-through effect, we used the 3D LAMINART model that has previously explained figure-ground percepts with static displays (Grossberg, 1997; Grossberg & Kelly, 1999; Kelly & Grossberg, 2000; Ross, Grossberg, & Mingolla, 2000). The 3D LAMINART model elaborates on the idea that visual processing involves two complementary pathways (Grossberg, 1997; Grossberg & Mingolla, 1985a, 1985b). One pathway (LGN → V2 Monocular Surface → V4 Binocular Surface) is called the Feature Contour System (FCS). It provides an explicit representation of visual surfaces, which includes filling-in of brightness and color information as defined by the complementary system. The other pathway (LGN → V1 Monocular Boundary→ V1 Binocular Boundary → V2 Binocular Boundary → V4 Binocular Surface) is called the Boundary Contour System (BCS). It codes oriented luminance edges and computes disparity differences between edges across the two eyes.

This model has been very successful at explaining and simulating properties of 3D depth perception; including explanations of how depth is derived from stereopsis (Grossberg & Howe, 2003; McLoughlin & Grossberg, 1998), how 3D surface perception is related to neon color spreading and transparency (Grossberg & Yazdanbakhsh, 2005), how depth percepts derive from da Vinci steropsis (Cao & Grossberg, 2005), how judgments of lightness depend on surface properties (Grossberg & Hong, 2006), and how 3D percepts derive from texture elements (Grossberg, Kuhlmann, & Mingolla, 2007).

In the following section we describe the 3D LAMINART model and explain how it responds to different types of static spatial stimuli. Equations for the simulations are described in the Appendix. In subsequent sections we show how the very same mechanisms, when analyzed through time, explain the appearance of shine-through. The model mechanisms are described in neurophysiological terms, as identified by Cao and Grossberg (2005).

2.1. Static stimuli and false disparity matches

It is well established that the visual system uses binocular disparity as a way of computing depth percepts. A critical issue faced by a system using binocular disparity is the existence of false binocular matches, where an element in one eye has a disparity match with a different element in the other eye. The existence of false binocular matches is sometimes described as the "correspondence problem," where the task is for the visual system to correctly identify how elements in one eye should be matched with elements in the other eye. The 3D LAMINART model provides a solution to the problems raised by false binocular disparity matches. Cao and Grossberg (2005) showed that the 3D LAMINART model matches human percepts for a variety of stimuli that vary in spatial layout, including a variety of cases that were studied empirically by McKee, Bravo, Taylor, and Legge (1994).

Fig. 2 schematizes the model components that are most important for explaining the appearance of shine-through. The drawn connections between different components skip many intermediate stages, although those stages were included in the quantitative simulations described below. Details of those intermediate steps, and their role in a variety of perceptual and neurophysiological phenomena, can be found in the Appendix and in Cao and Grossberg (2005).

Before turning to an analysis of shine-through it may help to introduce the model mechanisms by considering how they deal with depth computations for static stimuli. Fig. 3 summarizes a model simulation where false binocular matches are generated and subsequently removed.

The model simulations are reported in a consistent format. The horizontal arrow at the top indicates a progression through time, with the numbers below the arrow line indicating a moment in milliseconds after stimulus onset. The boxes under each number portray image planes of the different model stages in Fig. 2. The bottom two boxes schematize the stimulus drawn to the left and right eyes. The stimulus was deliberately chosen to resemble the stimuli used in the shine-through experiments, however, in Fig. 3 the vernier and non-vernier elements are presented simultaneously

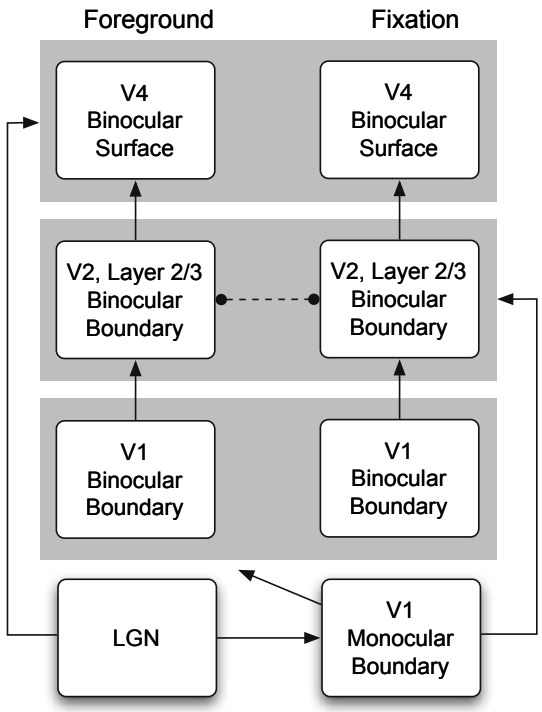

Fig. 2. A schematic of the major components of the 3D LAMINART theory. Each box corresponds to a retinotopic representation of visual space. Binocular representations include foreground and fixation planes.

(in the shine-through experiments the non-vernier elements temporally follow the vernier elements). The stimuli in Fig. 3 are identical across the two eyes, with no binocular disparity.

The frames in the second row from the bottom indicate the values of cells in the LGN (left) and V1 Monocular complex cells (right). Only image planes corresponding to the right eye are shown as the activities for the left eye would be identical for this simulation. Within every image plane the values are normalized so that the largest activity across the entire simulation is assigned the maximum white or black value. Except for the V4 Binocular Surface stages the value zero is assigned the middle gray value. All activities are gamma corrected ($\gamma = 0.45$) to allow small activity values to be visible. For the model stages that contain cells of different orientations (horizontal and vertical), activities from vertically tuned cells are assigned light gray and white values, while activities from horizontally tuned cells are assigned dark gray and black values.

The remaining boxes show image plane activities for model stages with different disparity planes. The third row from the bottom shows activities from binocular orientation sensitive cells of V1. A competitive circuit at this model stage insures that these cells obey an obligate property (Poggio, 1991), so that a cell remains active only when it receives input from both eyes that comes from cells coding the same orientation and luminance polarity at positions appropriate for the plane's disparity sensitivity. Cells at this stage of the model can respond to a true disparity between stimuli across the two eyes but can also respond to a false disparity match due to co-occurence of edges across the eyes that appear at the appropriate positions but are for different objects.

The fourth row from the bottom shows activities of binocular cells of Layer 2/3 in area V2. These orientation sensitive cells code boundaries at a specific disparity. These cells have a complex receptive field that samples inputs from the V1 Binocular cells at the same disparity as well as monocular inputs at the appropriate line-of-sight retinal positions. Line-of-sight refers to the position in 2-D space that a pixel occupies as a function of the 3D position implied by the disparity plane. In addition, the V2, Layer 2/3 cells are part of a cooperative–competitive network that groups together and enhances some boundaries while suppressing others. This cooperative–competitive network is described in more detail below. Finally, these cells are part of an additional competitive network that operates between disparity planes. Responses at corresponding positions compete across the disparity planes.

The top row shows activity across the V4 Binocular surfaces for each disparity. Here the boundaries across the V2, Layer 2/3 cells define closed surfaces. Inputs from the LGN monocular stages feed in to the V4 stage and spread brightness signals across the surface regions defined by the V2, Layer 2/3 boundaries. Activities across the V4 stages correspond to perceptual experience of brightness. Judgments about the target are derived from the properties of these activities.

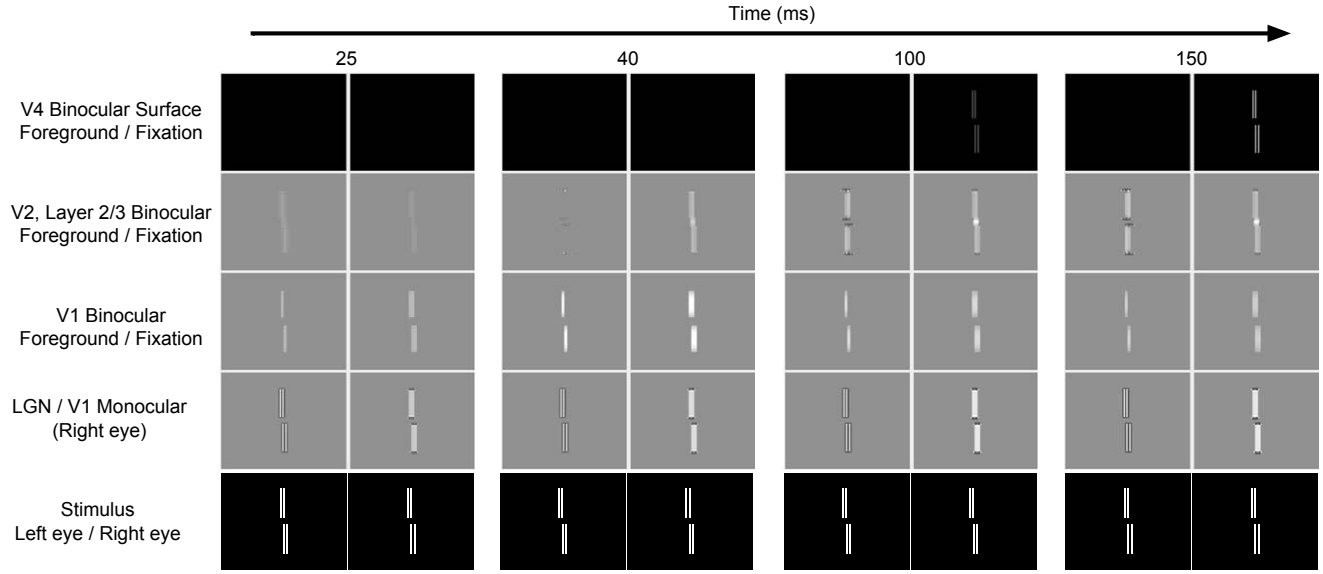

Fig. 3. Simulation results that demonstrate how the 3D LAMINART model removes false binocular matches. The stimuli for the left and right eyes are identical (see the text for details).

Although the stimulus is static during the simulation, the model processes unfold over time. At 25 ms after stimulus onset there are several key processes to understand. The V1 Monocular stage detects the edges of the vertical stimulus elements. These activities project to the V2, Layer 2/3 Binocular cells along their line-of-sight positions. The V1 Binocular cell responses are generated at both disparity planes. Cells on the fixation plane (right) are tuned to a disparity equal to zero. For these cells the elements in the left and right eyes always find their corresponding match in the other eye. Binocular matches are also found for the foreground disparity plane (left), whose cells are sensitive to a disparity of minus one. A binocular cell of this type at a specific pixel, i, looks for input from the left eye that is at a position one pixel to the right, $i + 1$, and for input from the right eye that is at a position one pixel to the left, $i - 1$. The far right elements (top and bottom) in the left eye are spatially located at a disparity relative to the far left elements (top and bottom) in the right eye. Such disparity differences can be interpreted as indicating that an element is closer than fixation at a spatial position between the position of the contributing elements. The V1 Binocular cells code boundaries with an allelotropic shift that spatially locates the matched edges at a position between the positions of the edges in the left and right eyes. The responses at the foreground V1 Binocular cells correspond to false binocular matches.

Thus, across the V1 cells, the model detects the monocular presence of edges and detects the presence of matches between the two eyes at the appropriate disparity. These two sets of signals project to the V2, Layer 2/3 stages. As can be seen in Fig. 3 at time 25, there are weak representations of the stimulus at both the foreground and fixation disparity planes of the V2, Layer 2/3 cells.

To remove the false binocular matches in the foreground disparity plane at the V2, Layer 2/3 stage, 3D LAMINART uses a line-of-sight competition between the binocular disparity planes of the V2, Layer 2/3 stage. The competition is biased in favor of information at the fixation plane. The effects of this competition are visible in the simulation at time 40. Here the false binocular matches across the V2, Layer 2/3 foreground stage have been inhibited by the signals across the fixation plane. With the continued presence of the stimulus, the activities across the network generally grow in strength. By time 100, the V2, Layer 2/3 cells in the fixation plane are strong enough to define closed regions, and brightness signals flow across the V4 fixation plane that correspond (veridically) to the stimulus at the fixation plane. Cell activities at the V2, Layer 2/3 foreground stage also grow in strength, but they are never able to form a closed region across the V4 foreground stage and so the stimulus is never represented at the foreground depth. Once established, this pattern of activity stabilizes, as evidenced by the pattern at time 150, which is essentially the same as at time 100.

The net result of the competition between V2, Layer 2/3 disparity planes is that the system is biased to encode boundaries at the fixation disparity plane, unless there is sufficient evidence that boundaries should be coded at another depth plane. This additional evidence can occur, for example, when there is a true binocular disparity shift between elements in the left and right eye. This effect is shown in Fig. 4, where the stimuli in the left and right eye are different with a negative disparity shift (the vernier in the left eye is slightly to the right of the vernier in the right eye). This stimulus is derived from the stimuli in Fig. 3 by removing from the left eye the far left elements from the top and bottom and removing from the right eye the far right elements. Here the fixation disparity plane has no binocular matches among the V1 Binocular cells.

Although there are no binocular matches at the fixation plane, the V2, Layer 2/3 binocular cells at the fixation disparity plane do continue to receive monocular inputs from their corresponding line-of-sight positions, but these activities are quickly inhibited by signals from the corresponding positions in the foreground disparity plane. The foreground V1 Binocular cells respond to the disparity across the eyes and this input combines with monocular input to allow the foreground V2, Layer 2/3 cells to win the competition over the fixation plane V2, Layer 2/3 cells, which receive only monocular input.

The model emphasizes that identifying binocular representations is less a task of matching elements across the eyes and more a matter of weighing evidence for different types of information. The interactions between disparity planes lead to a coherent choice for an object. It is this emphasis on competition between disparity planes that allows Cao and Grossberg (2005) to explain a variety of spatial effects on binocular vision.

Significantly, the same interactions can receive additional sources of information and can be influenced by a variety of spatial relationships. For example, Cao and Grossberg (2005) show how feedback from the V4 Binocular surface stages can provide

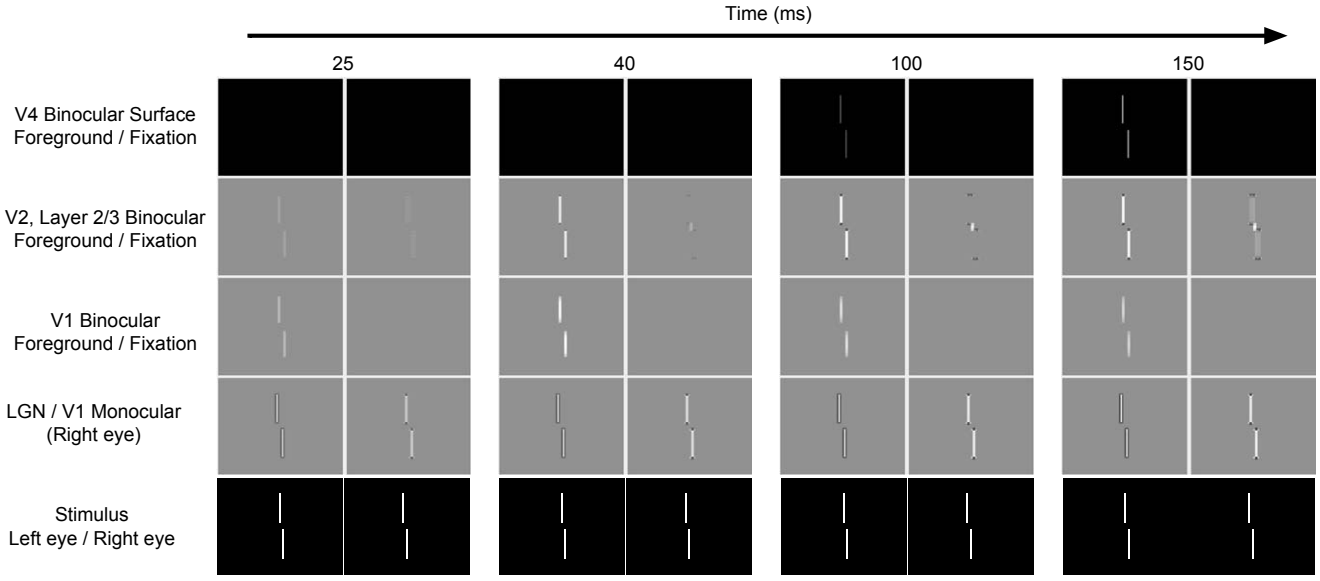

Fig. 4. Simulation results that demonstrate how the 3D LAMINART model correctly represents a stimulus that has a disparity shift between the left and right eye (see the text for details).

additional evidence that is used to determine which boundaries should win the disparity plane competition. Boundary arrangements that can support a visible brightness percept receive stronger feedback than boundary arrangements that do not match with the brightness inputs. This feedback also allows boundaries that support high contrast percepts to win the disparity competition against low contrast percepts. More generally, Cao and Grossberg (2005) argue that perceptual grouping of elements strengthens and weakens different boundary arrangements and thereby becomes an integral part of removing false binocular matches.

The current simulations do not include binocular cells that are sensitive to depths behind the fixation plane. Such cells would also produce false binocular matches to the stimuli shown in Fig. 3. However, these false binocular matches would be inhibited by the same competition between disparity planes. In addition, Cao and Grossberg (2005) hypothesized that nearer disparity planes tend to inhibit farther disparity planes more than the reverse. This asymmetry between near and far depths insures that when false binocular matches do produce visible percepts (as described below) it is amongst the near depth planes rather than the far depth planes.

It may seem that the network is overly complex, but every property of the model is included to address either neurophysiological, psychophysical, or phenomenological data. In this respect, the model is as simple as possible, considering the number of phenomena that it addresses. A full justification of every stage of the model would be beyond the scope of the present paper.

In the following analysis we analyze how the model mechanisms behave over time and demonstrate that the mechanisms identified as necessary for creating consistent 3D percepts by removing false binocular matches are also able to explain the properties of shine-through.

2.2. Responses to a dynamic target vernier

To analyze the dynamics of the 3D LAMINART model, we start with the case of a brief spatially offset vernier. This stimulus both helps to describe how the model behavior develops over time and also establishes how the model behavior is related to an empirical measurement. It also establishes a baseline of model behavior that can be compared against cases when the vernier is followed by a mask. For this and all subsequent simulations, the same stimulus is presented to the left and right eyes with no disparity difference. As a result, the stimulus can be identified by looking at the responses from the LGN and V1 Monocular Boundary responses for the right eye. The signals would be identical for the left eye. A model simulation is summarized in Fig. 5, where the target is presented for 20 ms with an onset at time zero. The target is so brief that much of the activity in the system develops after stimulus offset.

The model's dynamic behavior can be understood by exploring how the activities at each stage of the model vary through time. The LGN signals code the pixel values of the stimulus and grow progressively weaker at offset of the stimulus (time = 20 ms).

The V1 Monocular stage shows oriented responses that initially indicate the vertical and horizontal edges of the stimulus. The signals persist beyond the offset of the stimulus, but at times beyond 220 they show a weak orientation after-response, where cells sensitive to the orthogonal orientation respond. These after-responses are important for controlling visual persistence (Francis, Grossberg, & Mingolla, 1994) and also explain a variety motion (Kim & Francis, 2000), color (Francis & Rothmayer, 2003), and shape aftereffects (Francis & Grossberg, 1996). However, these after-responses do not play a major role in the simulations of shine-through, so they will not be discussed further.

The V1 Monocular signals from each eye feed in to the V1 Binocular cells. Because the left and right eyes view the same stimulus, the V1 Binocular cells code the stimulus at the fixation plane but not at the foreground disparity plane. As signals in the V1 Monocular stages weaken, the competitive circuits at this stage remove the persisting signals.

The V1 Binocular and the V1 Monocular signals feed in to the V2, Layer 2/3 stages. Cells in each V2, Layer 2/3 disparity plane receive signals from two sources: V1 Binocular cells of the same disparity and position, and V1 Monocular signals from both eyes at corresponding line-of-sight positions. The monocular signals are weak at the foreground stage because signals from the different eyes project to different line-of-sight positions (because the stimulus disparity does not match the required disparity for the foreground plane). In contrast, the fixation plane signals strongly indicate the position of the stimulus edges. The fixation plane cells receive matching inputs from the V1 Binocular cells and the V1 Monocular cells of each eye. The imbalance of activity in the foreground and fixation planes of the V2, Layer 2/3 cells is further augmented by line-of-sight competition between the disparity planes, which is biased to favor the fixation plane. Feedback from a cooperative–competitive loop (described below) allows signals at this stage to substantially outlast the inputs from the V1 Binocular and V1 Monocular stages (Francis et al., 1994).

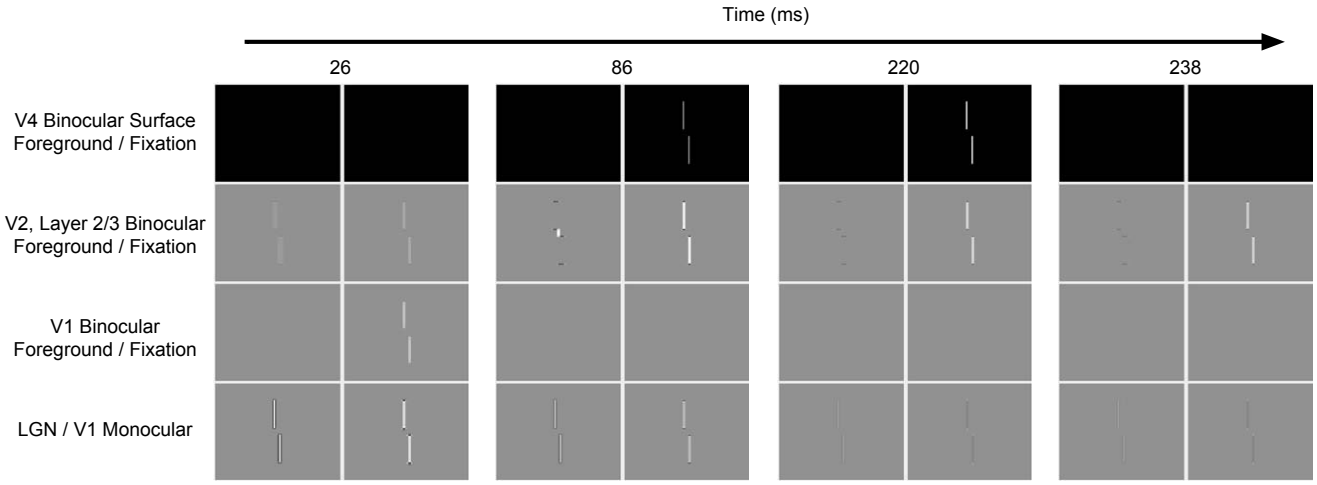

Fig. 5. Simulation results for a vernier target presented without a following mask. The V4 Binocular Surface stage only shows activity at the fixation plane (see the text for details).

The V4 Binocular Surface stages code the representation of perceived brightness. When the V2, Layer 2/3 cells become strong enough (time 86), they block the flow of brightness information at the V4 image plane of the same disparity. Here, the V2, Layer 2/3 cells in the fixation plane define a closed surface and brightness signals from LGN Monocular stages are trapped by the boundaries and grow in strength. This representation continues until time 238, when the horizontal fixation plane boundaries weaken so much that they no longer define a closed surface. When the contour breaks, the brightness information flows out and the representation of the target disappears from the filling-in stage.

Fig. 6a plots the evidence for the target vernier as a function of time. This evidence is computed as a difference of template matches for the vernier (with an offset the same as the target) and an anti-vernier (offset in the opposite direction). For each template, the sum of the V4 cell activities across a disparity plane was calculated. For a given pattern of cell activities, there may be some patterns consistent with both the vernier and the anti-vernier templates; and the difference indicates the advantage of one percept over the other. Further details of these calculations can be found in the Appendix. It takes around 50 ms for evidence for the vernier target to appear in the fixation plane. The evidence then grows stronger as the brightness signals at the V4 stage converge toward their resting values. The sudden drop in target evidence just before time 250 is due to the weakening of the V2 Layer 2/3 boundaries in the fixation plane. When these boundaries go below a threshold value, they no longer contain the brightness information about the target and the target disappears from the V4 stage. No evidence is found in the foreground plane because, as Fig. 5 shows, the boundaries of the foreground disparity are too weak to support filling in at the V4 stage.

3. Shine-through

The presentation of a grating mask after the target offset fundamentally changes the model's behavior. Fig. 7 shows a simulation where the mask consists of nine elements and is presented for 300 ms after target offset. The first important property in this simulation is evident at the LGN stage, where the target and mask elements are temporally integrated and signals coexist at time 26. As a result of these coexisting signals, the foreground V1 Binocular disparity sensitive cells find disparity matches for the edges of the target at time 86. These are essentially the same false matches analyzed in Fig. 3, where the binocular matches at the top of the image are the result of edges from the central mask bar in the left eye matching up with edges of the target bar in the right eye. Likewise, the binocular matches at the bottom of the image are due to edges from the central mask bar in the right eye matching up with edges of the target bar in the left eye. The spatial offset of the target vernier elements provide the spatial shifts that are detected by the disparity sensitive cells.

These false binocular matches combine with monocular signals at the V2, Layer 2/3 foreground stage. This combination of signals explains the smear of white vertical signals at the V2 Layer 2/3 foreground stage at time 86. Normally, these signals are filtered out by the competition across disparity planes in the V2, Layer 2/3 stage of the model. That this system normally works is evidenced by the absence of foreground signals at other positions corresponding to the mask elements. These mask elements do not produce any foreground binocular signals, but they do send monocular signals to the appropriate positions of the foreground V2, Layer 2/3 cells. Those monocular signals are inhibited by the presence of corresponding signals in the fixation plane, which contains both monocular and binocular inputs.

We will demonstrate that shine-through occurs when the competition between disparity planes experiences disinhibition. This weakening of the competition allows false binocular matches (generated by the target and mask) to produce boundaries at the foreground disparity plane.

The ineffectiveness of the disparity competition is due to two sources of inhibition within the fixation disparity plane. First, there is a spatial competition among boundaries of the same orientation and nearby positions. This spatial competition among the target and mask elements at the fixation plane initially favors the target because it was presented earlier and therefore has stronger boundaries than the mask. The spatial competition thus initially weakens the boundaries corresponding to the central element of the mask. As a result, the boundaries in the foreground that correspond to the line-of-sight locations of the target are inhibited by between disparity inhibition from the target boundaries in the fixation plane. In contrast, the positions in the foreground that correspond to the location of the mask's central element are less inhibited by the disparity competition. This pattern of disparity inhibition generates a weak anti-vernier with a displacement opposite to the target in the foreground plane, as shown at the foreground V4 Binocular Surface stage at time 86 in Fig. 7.

Because the mask follows the target in time and has a longer duration, its boundaries eventually dominate the target's boundaries. As the boundaries of the central element of the mask increase in strength, they send stronger inhibition to the corresponding positions in the foreground disparity plane. Eventually (time 123), these signals inhibit the anti-target representation in the foreground disparity plane.

The second necessary source of inhibition comes from a long-range grouping process performed by the V2, Layer 2/3 Binocular cells. A full description of this process is given in the Appendix, but the basic idea is that a set of horizontally tuned orientation sensitive cells group together when they are aligned horizontally, while a set of vertically tuned orientation sensitive cells group together when they are aligned vertically. This grouping initiates a feedback process that strengthens the activities of the cells in the group and can produce "illusory" boundaries between disconnected sets of boundaries. The grouping obeys a bipole property (Grossberg & Mingolla, 1985a, 1985b) that allows for interpolation of contours between supporting contours but prohibits extrapolation of contours in to empty space.

The grouping also introduces a competition among vertical and horizontal boundaries, such that a strong grouping of one orientation can prevent grouping of the other orientation if the first orientation crosses the potential path of the other grouping. This competition means that the overall geometric spatial layout of the scene can have a big influence on the way oriented boundaries in the scene group together. Grossberg and Mingolla (1985a, 1985b) called this competition "spatial impenetrability" because it prevented a grouped set of boundaries from penetrating across an even stronger grouping of boundaries of a different orientation.

For the 9 Elements mask in Fig. 7, the V2, Layer 2/3 fixation plane shows horizontal responses at the top and bottom of the mask that support the long-range grouping process that strengthens these horizontal boundaries. This grouping also prohibits long-range grouping processes that might otherwise support the vertical boundaries of the target and mask from developing within each element. The horizontal grouping effects can be seen in Fig. 7 at time 86 and later. The black horizontal lines at the top, middle, and bottom of the mask in the V2, Layer 2/3 fixation plane indicate the strong grouping of horizontal activities. The vertical (white) bars grow weaker over time as the horizontal (black) bars grow stronger because the horizontal groupings inhibit the vertical groupings.

For the target alone case (Fig. 5), the vertical boundaries of the target did group together because the few horizontal boundaries at the top and bottom of each vertical line were weak. This

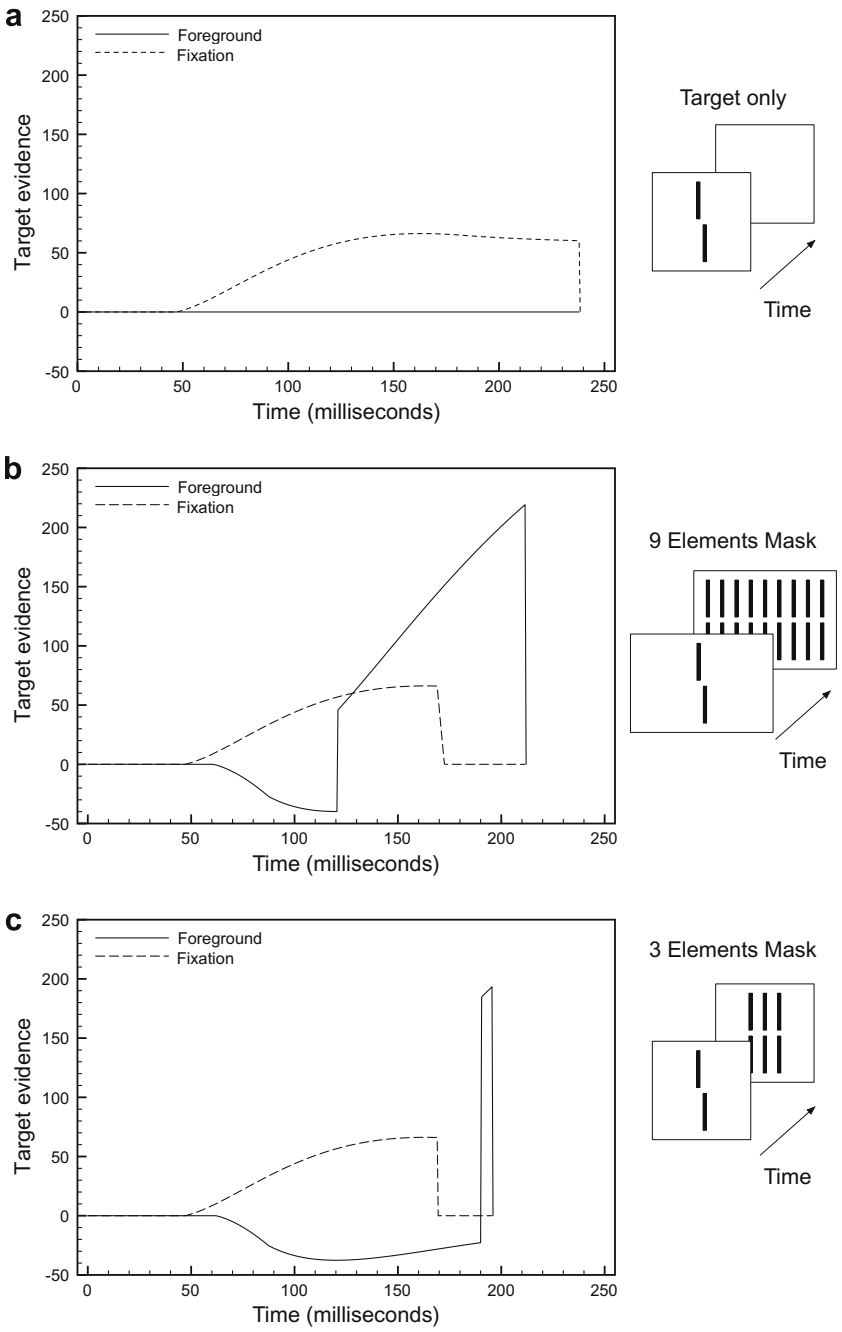

Fig. 6. Plots of target evidence as a function of simulated time. Separate curves show the match between a template of the target vernier and activity patterns at the foreground and fixation disparity planes. Positive values indicate evidence for a vernier shifted in the same direction as the target, while negative values indicate evidence for a vernier shifted in the opposite direction as the target. (a) When only the target is presented, all evidence is in the fixation plane. (b) A vernier target followed by a 9 Elements mask has more evidence in the foreground than in the fixation disparity plane. This simulation indicates shine-through. (c) With a 3 Elements mask, there is more evidence in the fixation plane than in the foreground. This simulation demonstrates masking rather than shine-through.

strengthening of vertical boundaries in the fixation plane provided additional inhibition to the foreground disparity plane. Since the 9 Elements mask removes the vertical grouping, the between disparity plane inhibition is weaker.

Fig. 7 shows that the net result of the spatial competition and the prevention of vertical grouping is that the vertical V2, Layer 2/3 boundaries generated by the target in the fixation plane are much weaker than if there were no mask. As a result, the boundaries corresponding to the false matches in the V2, Layer 2/3 foreground receive less inhibition than if there were no mask. Thus, at time 123, the false disparity matches generated by the target with the central mask bar have a strong enough representation at the foreground V2, Layer 2/3 disparity plane to define filling-in regions at the V4 Binocular Surface stage. This is the shine-through effect.

The shine-through effect has a fairly short duration because the boundaries introduced by the target fade away, as evidenced by the absence of false matches in the V1 Binocular foreground at time 123. As these signals weaken, the V2, Layer 2/3 boundaries at the same disparity also weaken and eventually no longer block the spreading of brightness information. The target percept at the V4 Binocular Surface disappears as soon as the brightness information spreads outside of the containing boundaries (time 211).

Fig. 6b plots the evidence of the target in the foreground and fixation plane V4 Binocular Surface stages. Evidence of the target

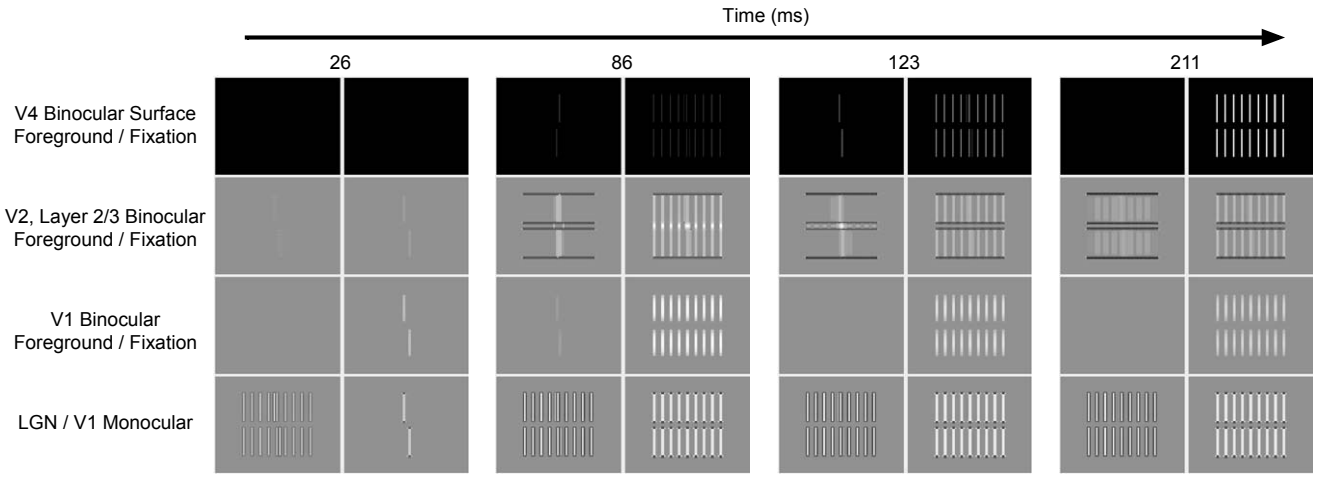

Fig. 7. Simulation results for a vernier target presented with a 9 Elements mask. Shine-through is indicated at time 123 with a representation of the vernier in the foreground V4 Bincoular Surface stage (see the text for details).

appears in the fixation plane before the spatial competition from the mask weakens the target boundaries there.

In the foreground disparity plane, the target evidence first indicates an anti-vernier (as shown at time 86 in Fig. 7). The anti-vernier evidence is short-lived because the increasing strength of the mask boundaries inhibit the contours that support this evidence. However the false binocular matches generated by the target and central mask element contours also grow in strength and introduce evidence of the target at the foreground plane. As a result, the foreground plane has an arrangement of boundaries that define a closed surface and support a filled-in percept of an offset vernier target at the foreground V4 Binocular stage.

The target evidence related to shine-through can be stronger than the target evidence generated by the target without a mask (compare the heights of the curves in Fig. 6a and b). This is because the V4 Binocular stage activities that correspond to shine-through draw upon LGN Monocular inputs from positions corresponding to both the target and the central element of the mask. For the target only simulation, the LGN activities fade after target offset, so the monocular inputs that feed in to the V4 Binocular filling-in stage are gradually weakening. The mask stimulus is presented for much longer than the target, so the LGN activities of the central mask element consistently provide strong inputs to the V4 Binocular filling-in stage.

3.1. Absence of shine-through

Not every kind of mask generates shine-through. Herzog et al. (2001) reported that shine-through does not occur when the mask consists of less than seven elements. The model explains this sensitivity in terms of perceptual grouping of the horizontal responses at the top and bottom of the mask. When the mask consists of few elements, there is insufficient activity to trigger the neural feedback that is responsible for grouping the horizontal boundaries of the mask elements.

Fig. 8 shows the model's behavior when the target is immediately followed by a 3 Elements mask. At time 26, the model behavior is similar to the behavior generated by the 9 Elements mask in Fig. 7. At time 54 the V1 Binocular foreground stage finds false matches between the target and the mask's central element. These are essentially the same as the false matches in Figs. 3 and 7. At time 54, the spatial competition across the V2, Layer 2/3 cells in the fixation plane also leads to an anti-vernier representation of boundaries at the foreground V2, Layer 2/3 disparity plane, but these are currently too weak to support filling-in.

As for the 9 Elements mask, the spatial competition among boundaries at the fixation plane produces disinhibition that effectively strengthens the false match boundaries at the foreground stage of the V2, Layer 2/3 cells (time 105). However, the 3 Elements mask differs in a substantive way from the 9 Elements mask. With only three elements in the mask, there is no boundary grouping of the horizontal boundaries at the top and bottom of the mask in the fixation disparity plane. As a result, the vertical target (and mask) boundaries of each element group together and strengthen the vertical boundaries at the fixation plane. The presence of this vertical grouping is indicated by the white signals connecting the top and bottom parts of the mask elements in the V2, Layer 2/3 fixation plane. This grouping is absent in the 9 Elements mask after time 123 (Fig. 7). The net result is that for the 3 Elements mask, there remains substantial inhibition from the fixation plane to the foreground plane. As a result, the false matches in the foreground stage are inhibited and only briefly become strong enough to contain brightness signals at the V4 Binocular Surface foreground stage.

Fig. 6c plots the evidence of the target as a function of time. The target makes an appearance in the fixation plane, but disappears as lateral inhibition from the mask weakens the target boundary signals. This behavior is almost identical to the 9 Elements mask. The foreground disparity plane shows a weak anti-vernier, with a shine-through effect appearing only just before all target boundaries collapse. The shine-through effect appears when the boundaries that supported the anti-vernier collapse thereby revealing the presence of signals that coded the target vernier. Overall, the foreground plane does not provide consistent evidence about the direction of the target vernier.

Thus, the model explains why shine-through appears for some masks but not for others. In summary, the model suggests that shine-through is due to a combination of factors:

- The rapid temporal stimulus sequence allows responses to the target and mask to form false binocular matches at the foreground disparity plane.
- Lateral inhibition between the boundaries of the target and mask at the fixation plane weaken each other and so lead to a reduction in the between disparity competition that would usually allow the fixation plane boundaries to inhibit the false binocular matches in the foreground.
- Grouping of the horizontal boundaries of the mask prevents grouping of the target and mask vertical boundaries in the fixation plane. Normally, the vertical grouping would strengthen the vertical boundaries at the fixation plane and thereby insure

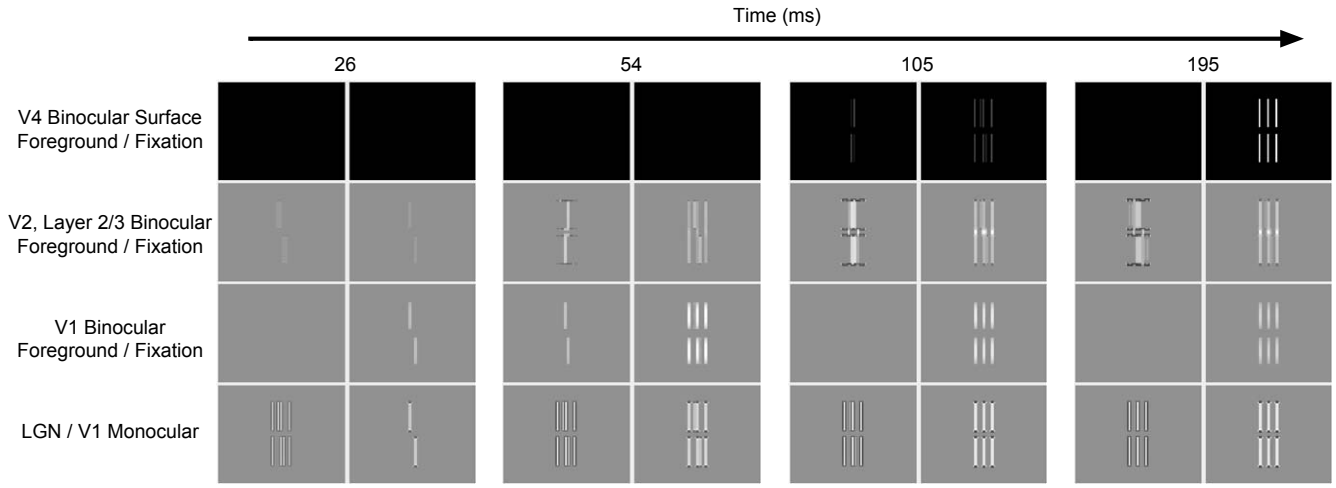

Fig. 8. Simulation results for a vernier target presented with a 3 Elements mask. The foreground V4 Binocular Surface stage has an ambiguous pattern of activity. Shine-through is not observed. See the text for details.

removal of the false matches at the foreground plane. The absence of this grouping leads to disinhibition of the false matches at the foreground V2, Layer 2/3 stage, which thereby support the shine-through percept.

- The boundaries of the false matches constrain monocular brightness inputs that come from both the target and the central bar of the mask.

All of these factors are necessary for shine-through to occur. Empirical work by Herzog and colleagues has focused on the importance of the perceptual organization of the mask elements for the appearance of shine-through. In the following simulations we show that the concept of perceptual organization corresponds to boundary grouping in the model and that such boundary grouping effects influence the appearance of shine-through in a way that closely matches the experimental data.

4. Variations in the number of mask elements

To quantify the shine-through effect in the model simulations, we measured the area under the target evidence curves in Fig. 6. We separately computed the integrated target evidence in the foreground and fixation disparity planes and took the larger integrated target energy as the model's interpretation of the visibility of the target. Similar simulations can be run for masks with different numbers of elements. The filled points in Fig. 9 plot the simulated target visibility as a function of the number of mask elements. The open symbols in Fig. 9 show experimental data for the very same stimuli and task (Herzog et al., 2001). For the experimental data, the measure of target visibility was percent correct identification of the target vernier offset direction. Both the simulation and the experimental data have the poorest target visibility for a mask of three to seven elements. For larger numbers of mask elements, the model predicts that shine-through occurs and the target visibility is high. The close relationship between the presence of shine-through and performance on the vernier detection task was also found by Herzog et al. (2001) (see their Table 1).

The model results match the pattern of the experimental data very well. Pearson's correlation between the simulated and experimental measures of target visibility is $r = 0.83$. The biggest discrepancy between the data and simulation is for a mask with one element. Here the simulation has a smaller integrated target evidence than the data would suggest. This difference might be because a single element mask introduces motion cues that can be used to judge the target vernier direction. Such effects have been reported with other types of masking stimuli (e.g., Ansorge, Breitmeyer, & Becker, 2007). The current simulations do not include motion calculations. If the mask with one element is not included, the correlation between the data and simulation results increases to $r = 0.97$.

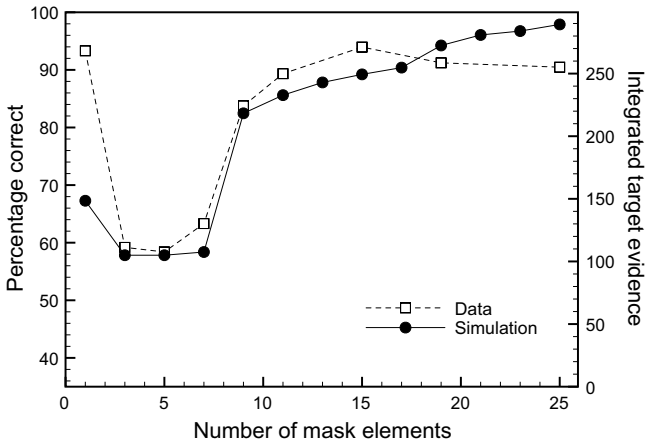

Fig. 9. The effect of the number of elements in the mask grating on target offset detection. For the simulation results, higher values of integrated target evidence occur with the presence of target vernier evidence in the foreground stage.

In general, the model predicts that whenever the horizontal boundaries at the top and bottom of mask elements group together to form a strong boundary, then these horizontal boundaries will prohibit strong vertical boundaries from forming at the locations of the target and the mask elements. The weaker vertical boundaries produce weak inhibition between the depth planes. This disinhibition allows the false matches from the V1 Binocular foreground to be represented at the V2, Layer 2/3 foreground stage, and these activities support brightness filling-in at the V4 Binocular surface stage. Masks with seven or fewer elements are unable to provide strong enough grouping to trigger this disinhibition.

Thus, any change in the spatial layout of the mask that prevents the formation of strong horizontal boundaries above or below the vertical boundaries of the target and mask elements should eliminate the shine-through effect. Without the horizontal boundaries, the vertical boundaries group together and inhibit the false matches at the foreground disparity plane.

5. Sensitivity to spatial grouping

The above analysis argues that shine-through occurs with a large number of mask elements because the additional elements produce contextual grouping effects that lead to disinhibition of false binocular matches in the foreground disparity plane. The model explanation quantifies the description proposed by Herzog et al. (2001), who couched their description in terms of image segmentation and perceptual grouping.

To demonstrate the effects of perceptual grouping, Herzog et al. (2001) varied the spatial layout of the mask grating. For all mask gratings, the central 5 Elements (called the kernel) were always present and unchanged. The size, location, or orientation of the other mask elements (called the context) varied. Thus, any local lateral inhibition due to the kernel would be unchanged across the different mask types, while long-range perceptual grouping effects could be modified by the context.

The left side of Fig. 10 schematizes five types of masks used by Herzog et al. (2001). The 25 Elements mask is like the mask used in Figs. 7 and 8, but it has 25 elements. In the length condition, the context elements are twice the length of the kernel elements. In the 90 degree tilt condition, context elements on each side of the kernel are rotated 90 degrees relative to the normal mask. In the gap condition, the elements directly on either side of the kernel are removed. The 5 Elements mask contains only the kernel. Empirical and simulated target visibilities for the 5 Elements and 25 Elements conditions have already been shown in Fig. 9.

The experimental data of Herzog et al. (2001) in Fig. 11 shows that only the 25 Elements mask is easily detected. In complementary reports, Herzog et al. (2001) found that target detectability was directly related to the appearance of the shine-through effect. The different perceptual reports are indicated in the data as good detection of the vernier offset for the 25 Elements mask but near guessing (50%) for the other masks. Not shown in Fig. 11 is a condition that Herzog et al. (2001) called 5 degree Tilt, because the current simulation cannot consider elements that are not oriented horizontally or vertically.

The model explains the effect of the different masks by how they influence the grouping of the boundaries. The right side of Fig. 10 shows the fixation plane V2, Layer 2/3 boundaries generated in response to the target and mask stimuli. For the 25 Elements mask, the horizontal boundaries group together, and this grouping prevents the vertical boundaries of the target and mask elements from grouping together. This leads to a weakening of vertical boundaries at the fixation plane and allows for shine-through to occur at the foreground disparity plane.

For the Length condition, the horizontal boundaries of the context cannot join with the horizontal boundaries of the kernel because they do not line up on the same horizontal row. As a

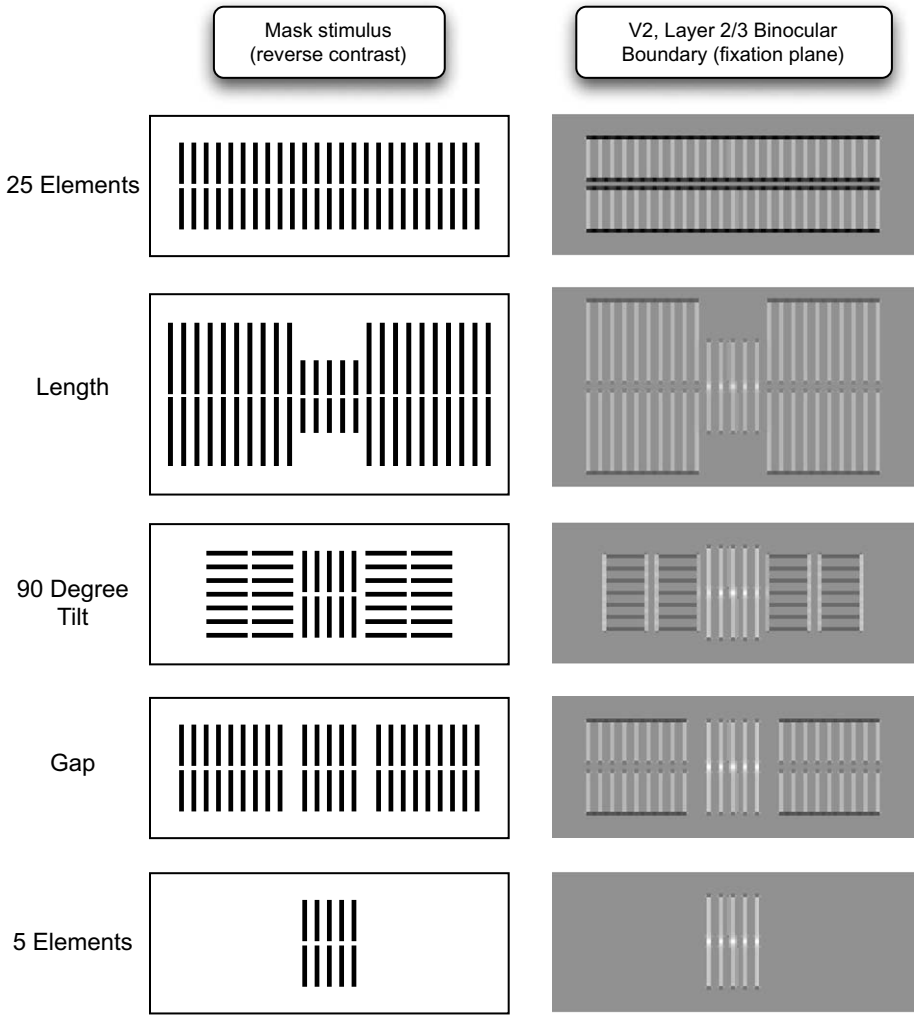

Fig. 10. Schematics and simulation results for various types of masks. The left images show reverse contrast schematics of the different types of masks. The image to the right of each drawn mask shows the horizontal (black) and vertical (white) activities of orientation sensitive cells in the fixation plane V2, Layer 2/3 cells. Only the 25 Elements mask groups the horizontal boundaries of the mask kernel (middle 5 elements).

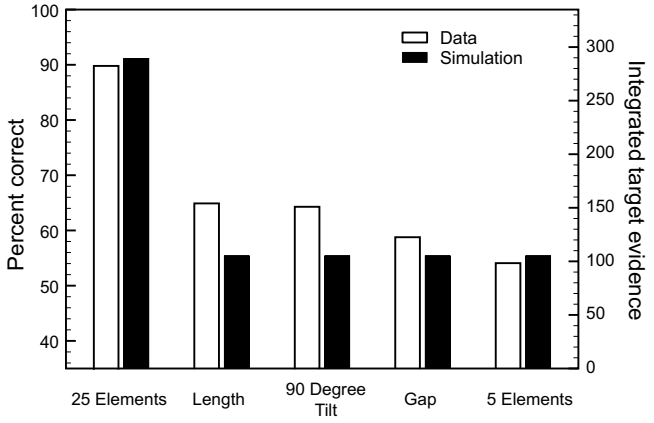

Fig. 11. Visibility of the target vernier for different types of masks. Both the empirical data and simulation results show the same pattern. Only the 25 Elements mask allows for good detection. This mask is also the only one to produce shine-through.

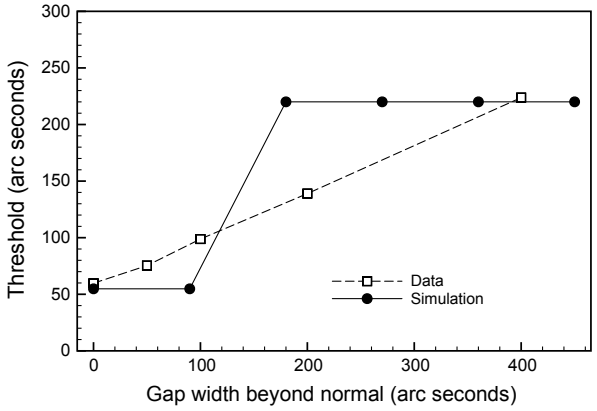

Fig. 12. Empirical and simulated vernier thresholds as a function of the width of the gap between the central 5 Elements of the mask and the context elements. Thresholds increase with gap width.

result, the kernel elements are isolated and do not produce strong horizontal grouping and so do not prevent the vertical boundaries of each target and mask element in the kernel from grouping together. Such grouping of the vertical boundaries strengthens the inhibition sent to the false matches in the foreground disparity plane and prevents shine-through.

The horizontal context elements of the 90 degree Tilt condition do not group with the horizontal boundaries of the kernel. For the specific stimuli we used, the horizontal contours from the rotated lines do not line up perfectly with the horizontal boundaries at the top of the vertical elements. Moreover, vertical responses at the ends of the rotated lines group together to form strong vertical contours that subsequently weaken any groupings generated by the horizontal rotated lines. The net result is that the kernel is again isolated and the vertical boundaries of each element group together and their increased strength prevents the appearance of shine-through.

The Gap condition produces a similar effect. The mechanisms responsible for joining disparate horizontal boundaries are unable to bridge the gap between the kernel and the context elements. As a result, the 5 elements of the kernel are functionally equivalent to the 5 Element mask by itself, which has too few elements to produce strong grouping and so does not produce shine-through.

The 5 Elements mask is similar to the 3 Elements mask previously discussed. There are too few elements to produce strong grouping among the horizontal boundaries of the mask, so the vertical boundaries of each element can group together and their increased strength prevents shine-through.

As Fig. 11 shows, in every case, the model's behavior matches the experimental data. Changes to the spatial properties of the mask that are quite far from the location of the target can have a profound influence on the appearance of shine-through and target detectability.

6. Variations in the gap

Herzog et al. (2001) studied the transition between shine-through and the absence of shine-through with variations of the gap display. In these experiments they measured a threshold offset of the target vernier rather than percentage correct detections of the target offset. Larger threshold values indicate that the target is more difficult to see, and small thresholds for these displays indicate shine-through (as verified in other experiments).

In one experiment, Herzog et al. (2001) varied the width of the gap. The open symbols in Fig. 12 show the experimental data from Herzog et al. (2001). A gap width of zero corresponds to the 23 Elements mask in Fig. 9. A gap width of 200 arc sec is similar to the gap condition in Fig. 10. The empirical data show a monotonic linear effect of increasing gap width on target threshold. Thus, with small gap widths the shine-through effect occurs and thresholds are low, while shine-through disappears (or is less common) for larger gap widths.

The model computes integrated target evidence, v, with larger values indicating better detection of the target. To convert these values in to thresholds, we computed a threshold as

$$T = 15 + \frac{335}{1 + \exp[a(v - b)]}. \quad (1)$$

The equation comes from Hermens et al. (2008) who used it to map target evidence to thresholds. We set parameters a and b so that an integrated target evidence of $v = 105$ mapped on to a threshold of $T = 220$ and $v = 290$ mapped on to $T = 52$. The value 15 reflects the empirical finding that thresholds almost never go below 15 arc sec for most observers. The value 335 establishes an empirical upper value for thresholds at 350 arc sec. For all the cases where we consider empirical data measured in thresholds, the model has big quantitative differences between cases with shine-through (large integrated target evidence) and cases without shine-through (much lower integrated target evidence). As a result, the parameters of the mapping play no role in the fundamental explanation of shine-through, but only insure that the model values fit within the range of thresholds that are found empirically.

As the filled symbols in Fig. 12 show, the model demonstrates an effect of increasing gap width, but it is a step function rather than a linear relationship. For small gap widths, the boundary completion mechanisms can connect the horizontal boundaries of the kernel and the surrounding context and produce strong grouping that leads to shine-through and small thresholds. When the gap width exceeds a hard threshold, the boundary completion mechanisms cannot connect the horizontal boundaries of the kernel and surrounding context. As a result, beyond that threshold no boundary grouping occurs and so shine-through does not occur and thresholds are high.

While the quantitative fit between the model and the data is not as good as for the other cases we will consider, a better fit could be generated by averaging results across several models with slightly different parameters. The empirical data in Fig. 12 are averaged across four observers, and there are variations among observers for whether shine-through is seen or not for a particular stimulus (see Table 2 in Herzog et al., 2001). The model predicts that this averaging is responsible for the linear shape of the curve.

In a second study of the influence of the gap on perceptual grouping and shine-through, Herzog et al. (2001) introduced an element in each of the gap positions and varied the element's size. Each gap element was like the other mask elements, except that part of the top and bottom of the gap element was erased or extended (see the stimulus schematic in Fig. 13a).

The open symbols in Fig. 13b show that empirical thresholds are lowest when the gap element is the same size as the other mask elements (length difference of zero). This corresponds to the 25 Elements mask grating in Figs. 10 and 11 and produces a striking shine-through percept. As the gap element size deviates from the normal mask elements (either bigger or smaller), the thresholds increase, which indicates that shine-through is less common.

The model explains these findings by the effect the gap element size has on the grouping of the horizontal boundaries of the mask. When the gap element is smaller than the other mask elements, the situation is analogous to the gap condition, where the grouping mechanisms cannot connect the kernel to the context elements. When the gap element is larger than the other mask elements, the situation is analogous to the length condition in Fig. 10.

It might seem that the horizontal boundaries along the middle of the mask should group together, since the gap element contributes a horizontal boundary that can connect boundaries together. Indeed, the geometry of the horizontal and vertical boundaries in the middle of the mask could allow for either a grouping of horizontal boundaries or a grouping of vertical boundaries. As indicated in the simulation activities of the V2, Layer 2/3 fixation plane cells in Fig. 13a, the vertical boundaries win the competition. The vertical boundaries win because there are more vertical boundaries along the pixels of a vertical line than there are horizontal boundaries along the ends of the different mask elements. Once established with extra strength from the grouping feedback, the vertical boundaries of the gap element prevent the grouping of the horizontal boundaries.

The filled symbols in Fig. 13b show the model's behavior as the size of the gap element changes. While a good fit is observed, the function is much sharper around zero than the actual data. This discrepancy is due to the current simulation of the model including only horizontal and vertical orientations. In an elaborated simulation of the model with more orientations, the horizontal boundaries of a gap element that was only slightly smaller or larger than the other mask elements would group with the rest of the mask. While grouping among nearby orientations is part of the general theory (e.g., Grossberg & Mingolla, 1985a, 1985b), it would require including many additional orientations in the current simulations, and each additional orientation would substantially increase the simulation time. Even without these additional model components, the current simulations provide a fairly good fit to the data.

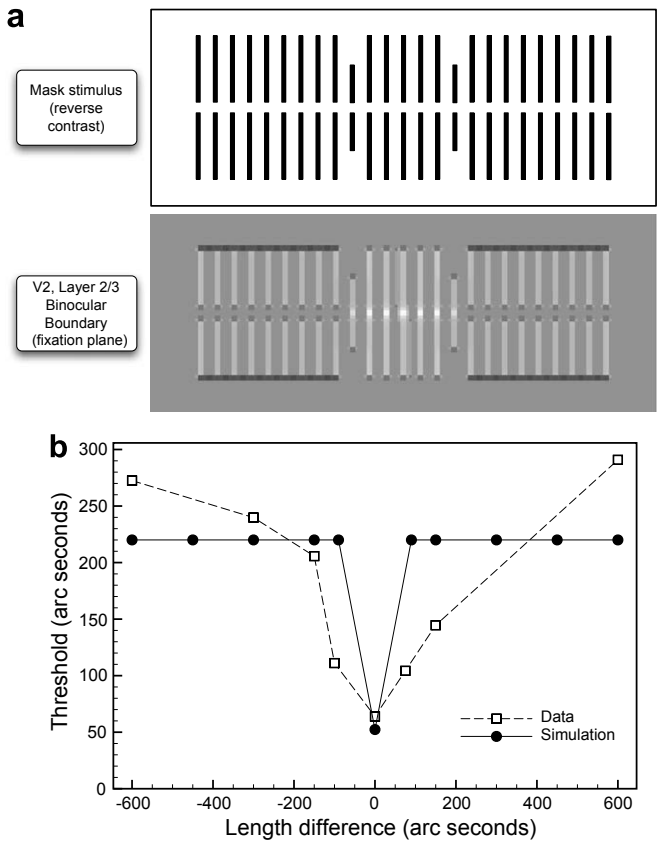

Fig. 13. Effects of gap element size on target offset thresholds. (a) Shows a schematic of the mask and the pattern of horizontal (black) and vertical (white) boundary responses at the V2, Layer 2/3 fixation plane. The gap at the top and bottom prevents the middle horizontal boundaries from grouping together. (b) In both the empirical and simulated results, thresholds are low only when the gap element is the same size as the remaining elements of the mask.

7. Manipulation of perceptual grouping

Herzog and Fahle (2002) used the shine-through effect to study properties of perceptual grouping. They showed that shine-through appeared when the mask elements grouped together to produce an independent coherent object. They demonstrated the importance of long-range perceptual grouping by considering displays with a 25 Elements grating and additional elements. These additional elements, though sometimes quite small, could prevent the appearance of shine-through, which was quite visible for the standard 25 Elements grating. The left side of Fig. 14 schematizes the various mask stimuli, while the right side shows the activities at the fixation disparity plane for the V2 Layer 2/3 cells at a moment in the simulation when the grouping of boundaries have nearly stabilized.

Herzog and Fahle (2002) used a standard condition that was the same as the 25 Elements condition in Figs. 10 and 11, so it is not shown in Fig. 14. For this mask, the horizontal boundaries of the mask elements group together. This horizontal grouping prevents strong grouping of the vertical boundaries of the target and the mask elements. Without such grouping, the inhibition from the fixation disparity plane to the foreground disparity plane is relatively weak and shine-through occurs.

In the 4 Lines condition, shown in Fig. 14, the vertical lines above and below the 25 Elements mask re-organize the boundary grouping. The new vertical boundaries group with the vertical boundaries of the 25 Elements mask between them. This vertical grouping prevents the horizontal boundaries from grouping together, and is conceptually similar to a long gap element, as in Fig. 13. Thus, no shine-through occurs.

In the 25 Lines condition, vertical lines are placed above and below each of the elements in the standard mask. The additional boundaries change the organization produced by the 4 Lines condition. Now, the horizontal boundaries of the additional lines group together and prevent the vertical boundaries from grouping together. This allows the horizontal boundaries of the original mask elements to group together and thereby produce a shine-through effect.

In the 4 Lines 90 condition, the four lines of the 4 Lines condition are rotated 90 degrees. The resulting horizontal boundaries do not group with the vertical boundaries of the original mask. As a result, the horizontal boundaries of the mask group together. This case is almost identical to the standard 25 Elements condition.

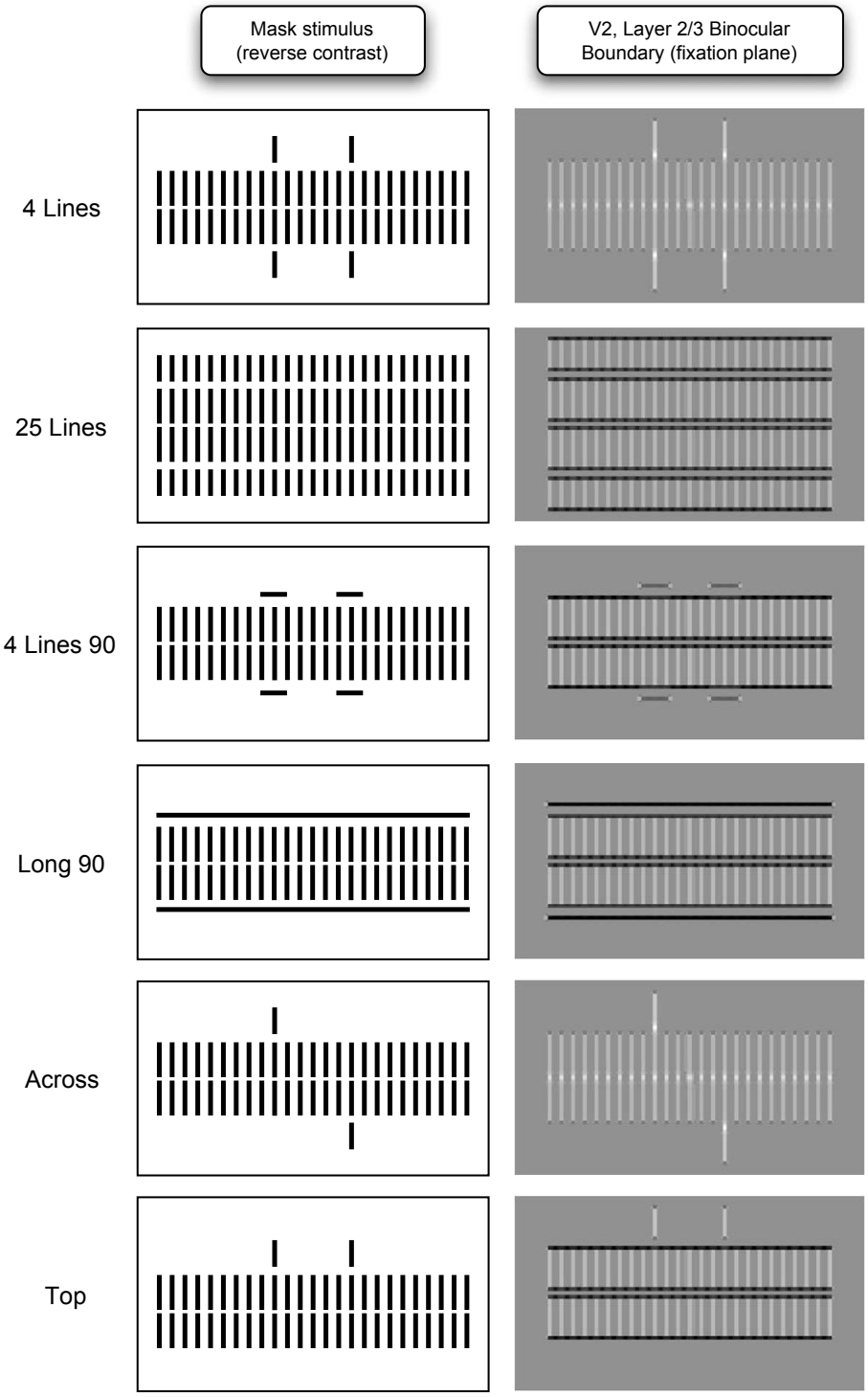

Fig. 14. The masks considered by Herzog and Fahle (2002) on the left and the horizontal (black) and vertical (white) boundaries at the V2, Layer 2/3 fixation plane. Only the 4 Lines and Across mask conditions do not allow for horizontal boundary grouping (see the text for details).

In the Long 90 condition, the horizontal lines are made long enough to cover all of the standard mask. The horizontal boundaries generated by these lines function in much the same way as the grouping of the horizontal boundaries in the standard 25 Elements mask.

In the Across condition, two of the lines from the 4 Lines condition are placed on opposite diagonal corners. These two lines function much the same as the lines in the 4 Lines condition. Namely, the vertical boundaries from the additional lines group with the vertical boundaries of elements in the standard mask. These vertical groupings prevent the horizontal boundaries along the top and bottom of the standard mask elements from grouping, and as a result the vertical boundaries of the standard mask can group together. The strengthened vertical boundaries prevent the appearance of shine-through by inhibiting the false matches in the foreground stage.

In the Top condition, two of the lines from the 4 Lines condition are placed on the top, but not the bottom, of the 25 Elements mask.

It might seem that this condition would be very similar to the Across condition since an equal number of vertical boundaries are available to group together. However, the global geometry of the scene leads to a very different arrangement of boundary grouping. The horizontal boundaries along the bottom of the mask are free to group together because there are no vertical boundaries that cut across the horizontal boundaries. Once this horizontal grouping is established, it weakens the grouping of the vertical boundaries above it. Without the grouping of vertical boundaries, the horizontal boundaries at the top and middle of the standard mask form horizontal groupings as well. The net result is that the weakened vertical boundaries lead to shine-through at the foreground disparity plane.

As shown in Fig. 15 for the Top condition, the grouping processes are dynamic and change over time. At 55 ms after target onset, both vertical and horizontal boundaries are weakly grouped together. As the boundary groupings grow in strength, they impose inhibition on the orthogonal orientation. At 68 ms, the vertical boundaries at the top of the mask have inhibited the horizontal boundaries that would otherwise form across the top of the mask. This inhibition does not last, however, because horizontal boundaries across the bottom of the mask do group together and start to inhibit the vertical boundary groupings. At 76 ms, the vertical boundaries have been inhibited and their weakening allows the horizontal boundaries across the top of the mask to group together again. By 88 ms the process has stabilized, with the horizontal boundaries dominating the grouping process.

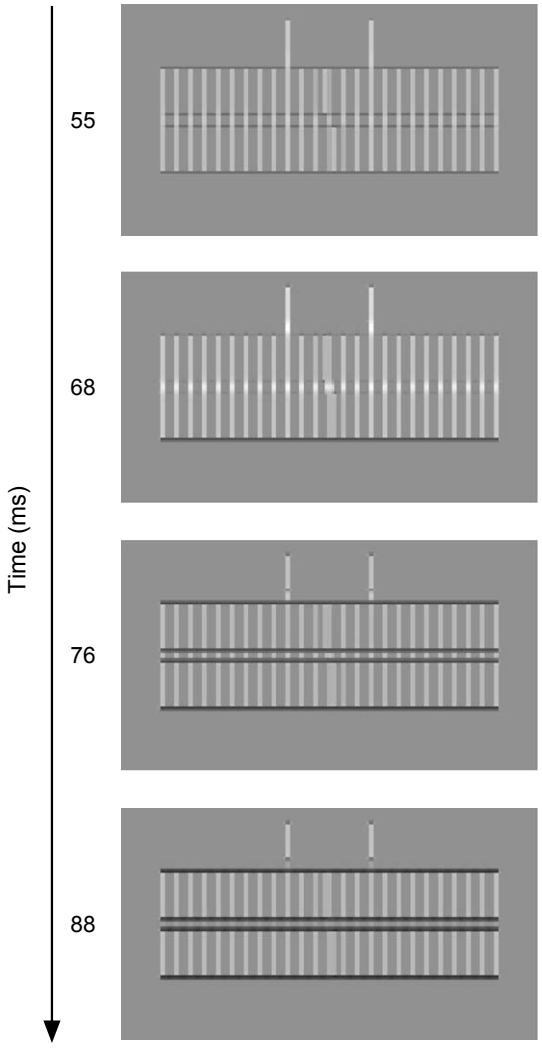

Fig. 15. The dynamics of fixation plane V2, Layer 2/3 boundary grouping for the Top mask. At time 68 the vertical elements group with the other vertical boundaries of the mask and inhibit the horizontal boundaries along the top. At the same time, the horizontal boundaries along the bottom of the mask group together. By time 76, the horizontal boundaries along the bottom of the mask inhibit the grouping of vertical boundaries. By time 88 the vertical boundaries no longer group and this allows all the horizontal boundaries of the standard mask to form groupings.

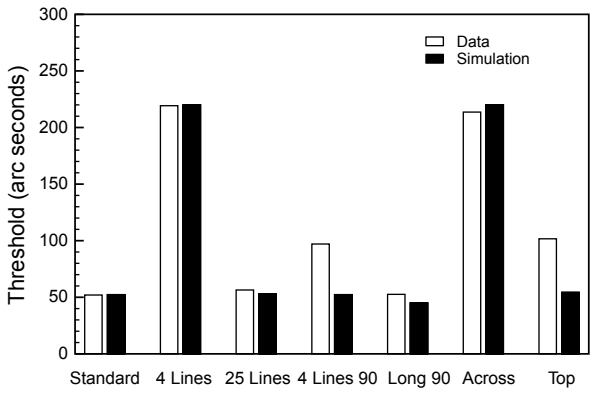

Fig. 16. Empirical and simulated thresholds for various types of masks. Only the 4 Lines and Across conditions do not show evidence of shine-through.

Fig. 16 shows that the thresholds created by the model simulation are very similar to the data from Herzog and Fahle (2002). Thresholds are low for the standard, 25 Lines, 4 Lines 90, Long 90, and Top conditions. Under these conditions, the geometry of the mask produces strong horizontal grouping of boundaries across the top and bottom of the mask elements. This grouping prohibits the formation of vertical grouping of boundaries that would strongly inhibit the false matches at the foreground disparity plane. This disinhibition leads to the shine-through effect and low thresholds. In contrast, the 4 Lines and Across conditions produce vertical boundary groupings that prevent the formation of the horizontal boundary groupings. As a result, shine-through is not generated for these conditions and the thresholds are higher.

8. Comparison to other models

Most models of spatial vision that deal with figure-ground distinctions (e.g., Craft et al., 2007; Geisler & Super, 2000) lack a description of temporal properties and so cannot address properties of the shine-through effect. Likewise, most models of temporal vision (e.g., Di Lollo, Enns, & Rensink, 2000; Francis, 2003; Weisstein, 1972) do not have a sufficient representation of the spatial properties of stimuli that would allow them to address the properties of the shine-through effect. The discovery of the shine-through effect has awakened the field to this shortcoming of current models of visual dynamics (Francis, 2007; Herzog, 2007).

As a result, there is currently only one alternative model that has been used to explain the properties of shine-through. The model exists in two versions. A one-dimensional version of the model (Herzog, Ernst, Etzold, & Eurich, 2003) was used to explain why increasing numbers of mask elements lead to a strengthening of the target percept. This one-dimensional model could not, of course, explain the influence of contextual elements as in Fig. 14. Very recently, this model has been extended to two dimensions (Hermens et al., 2008), and it accounts for many of the properties of shine-through discussed in this paper. While they differ in details and parameters, the two versions of this alternate model operate with the same basic principles. Both model versions differ substantially from the 3D LAMINART model.

Fig. 17 schematizes the model. It consists of two layers of neurons whose activities are described by Wilson–Cowan equations

for excitation and inhibition (Wilson & Cowan, 1973). The spatial kernels of these equations are set so that each cell receives recurrent excitation from spatial neighbors and recurrent inhibition from a larger set of cells around the excitatory set of cells. Thus, the equations describe a network of neurons whose connections define a recurrent center-surround receptive field for each cell. Although it risks grouping together systems that may have substantial differences, we will refer to this kind of system as a Wilson–Cowan Type Model (WCTM). Although there are many different ways to structure these kinds of models, the properties involved in explaining shine-through are fairly robust.

To explain the shine-through effect, the WCTM relies on the fact that the center-surround organization of the receptive fields cause the cells in the network to function as "irregularity" detectors. Consider the response of the inhibitory layer to mask gratings. Fig. 18 shows the pattern of activities across the cells in the inhibitory layer 40 ms after onset of the mask (the author thanks Frouke Hermens for providing the code to produce these simulations). For both the 5 Elements mask and the 25 Elements mask, the center-surround processing generates the strongest responses at the corners and outer edges of the mask. Activity is suppressed in the middle of each mask because of the strong inhibition from neighboring pixels. Pixels at the corners and edges of the mask receive less inhibition.

The pattern of activity schematized in Fig. 18 projects to the excitatory layer where cells receive those inputs with another center-surround receptive field. The inhibition at a cell in the excitatory layer will only be strong if it samples inputs from active cells in the inhibitory layer. This means that cells at the excitatory layer only receive strong inhibition if they are at pixels near the extreme edges of the mask. Since the vernier target is always at the middle of the image, the 5 Elements mask will send stronger inhibition to the pixels coding the target than the 25 Elements mask.

This property also explains why a gap in the mask leads to stronger masking. Pixels around the gap can have strong activity in the inhibitory layer because the gap does not inhibit nearby elements. These strong responses are then able to inhibit the pixels related to the target when these signals project to the excitatory layer.

Fig. 18 shows the model's response when only the mask is presented. The behavior is a bit more complex when the target precedes the mask because the target can inhibit some of the mask elements and thereby change the pattern of inhibition. In addition, the recurrent interactions between the excitatory and inhibitory layers can change the pattern of inhibition. These effects are generally small, however, and the basic properties that explain the relative strength of inhibition are as described above.

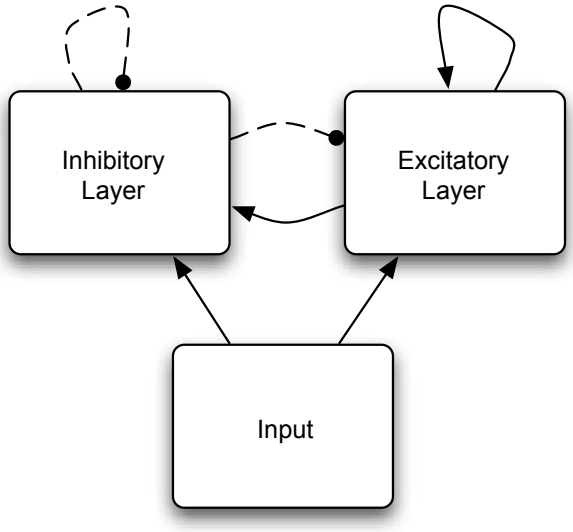

Fig. 17. A schematic of the Wilson–Cowan neural network that was used by Hermens et al. (2008) to explain the properties of shine-through.

Given its simplicity, the WCTM's breadth of data coverage is impressive. In addition to accounting for the effects of the number elements, the size of the gap, and the size of the gap element, Hermens et al. (2008) showed that the model accounts for several types of traditional backward masking effects and common onset masking. The model is less successful, however, in accounting for some of the context effects discussed above. In particular, the model's behavior does not usually match the empirical results when some of the mask elements are longer than the standard grating. As described above, the 3D LAMINART model provides a detailed explanation of these effects.

Perhaps more important than data coverage, is a consideration of how well the model deals with the phenomenological aspects of the shine-through effect. The WCTM correctly simulates that the target with a 25 Elements grating is more visible than the target with a 5 Elements grating. However, the model fails to explain why the target is perceived to occlude the mask for the 25 Elements grating. That is, improved visibility could occur with a perceptual experience of the target being perceived as part of the mask, or by inhibiting some part of the mask. The WCTM does not make a clear statement as to what the target percept should look like. The WCTM, of course, does not have an explanation for any figure-ground or occlusion distinctions, it only describes activities of model neurons. As a result, it fundamentally cannot explain one of the key perceptual aspects of the shine-through phenomenon.

More generally, the WCTM is currently disconnected from other visual phenomena. It is not a general model of visual perception, so even if it is successful at matching the empirical data on shine-through it will have limited utility in relating those data sets to other aspects of visual perception. This limitation is in sharp contrast to the 3D LAMINART model, which connects the mechanisms responsible for shine-through to many other visual phenomena, including: 3D perception (Grossberg & Howe, 2003; McLoughlin & Grossberg, 1998), figure-ground distinctions (Grossberg, 1997), illusory contours (Grossberg & Mingolla, 1985b), line motion (Baloch & Grossberg, 1997), brightness perception (Grossberg & Hong, 2006), aftereffects (Francis & Rothmayer, 2003), texture segmentation (Grossberg et al., 2007), visual persistence (Francis et al., 1994), metacontrast masking (Francis, 1997), and many others. The very same model mechanisms are involved in 3D LAMINART's explanation of all of these phenomena.

9. Predictions

We identified several novel masks that highlight the 3D LAMINART model's explanation of shine-through. We identified three cases where shine-through should not occur, and we identified two novel mask configurations that the model predicts should produce a shine-through effect. The left side of Fig. 19 shows a schematic of the masks. Fig. 20a shows the predicted integrated target evidence for each of the conditions. Fig. 20b converts the target evidence in to vernier offset thresholds using Eq. (1). The new masks differ from previously described masks in that the top and bottom context elements almost abut the top and bottom of the standard mask elements. All of the elements are positioned so that either a standard mask element or a top and bottom context element are shown, but never both in the same column.

The right side of Fig. 19 shows the activities at the fixation disparity plane of the V2 Layer 2/3 cells at a moment in the simulation when the grouping of boundaries has nearly stabilized. In all of

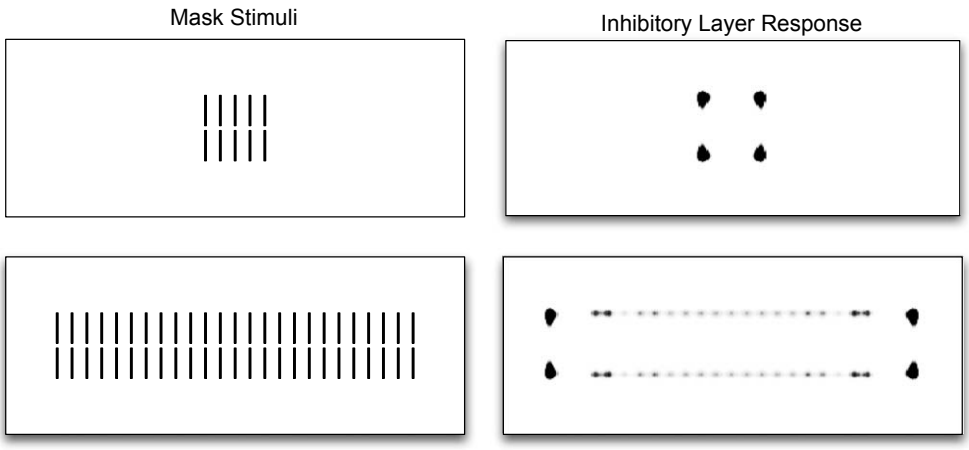

Fig. 18. Simulation activities (right) for the Wilson–Cowan type model of Hermens et al. (2008) for the 5 Elements mask and the 25 Elements mask stimuli (left). These signals inhibit nearby pixels at the excitatory layer. Since the target is at the middle of the display, the inhibitory signals for the 5 Elements mask will be stronger than the inhibitory signals for the 25 Elements mask.

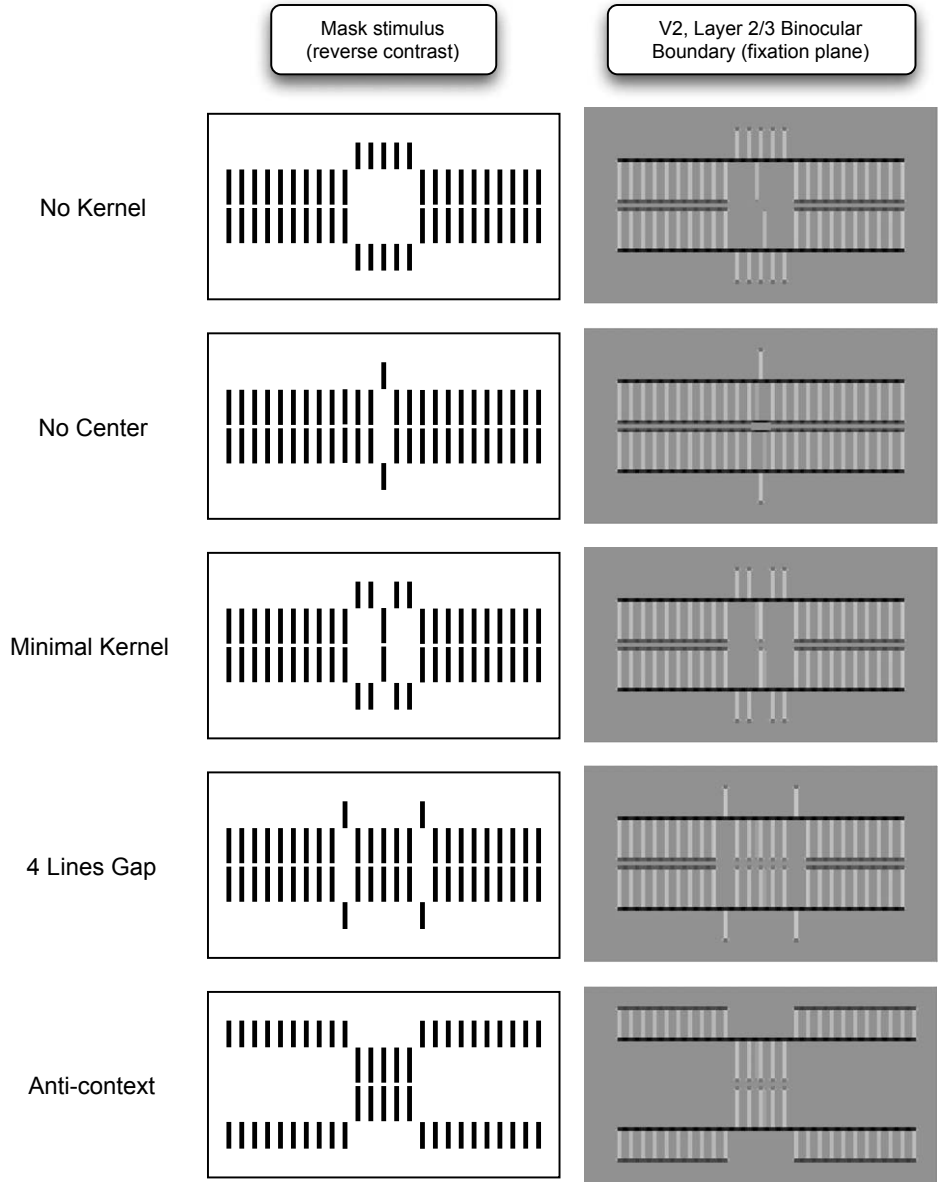

Fig. 19. Five new types of masks (left) and the response of the model at the fixation plane V2, Layer 2/3 cells. Because the top and bottom context elements abut the standard mask elements, all of the masks generate strong horizontal groupings.

these masks, the horizontal boundaries group together. This occurs because the horizontal boundaries generated by the top and bottom context elements can group with the horizontal boundaries of the standard mask elements. When the horizontal boundaries group together, they inhibit the vertical boundaries of the target and mask elements. However, grouping by itself is not enough to generate the shine-through effect.

The No Kernel condition removes the middle 5 Elements of the mask. The model predicts that this condition does not produce shine-through for two reasons. First, there is no opportunity for false binocular matches to be generated between the target boundaries and the central element of the mask. Second, without the surrounding elements of the mask, the target boundaries receive relatively little spatial competition. Thus, although the target's vertical boundaries do not get excitatory feedback from the grouping process, they are strong enough to inhibit any false matches in the foreground disparity plane. As Fig. 20a shows, all the integrated target evidence for the No Kernel condition is in the fixation disparity plane. The target evidence is fairly high because it receives little spatial inhibition; however, this stimulus sequence does not produce shine-through in the model.

The No Center condition introduces four of the 5 middle elements, leaving out only the central element of the mask. This central element is replaced by abutting elements above and below the standard mask elements. The abutting elements allow the horizontal boundaries of the mask to group together and prevent vertical boundaries from grouping. The elements around the target element introduce fairly strong spatial competition. However, it is the central mask element that introduces the false binocular matches that are critical for the appearance of shine-through. As Fig. 20a shows, the model predicts that this condition does not produce shine-through. All of the target evidence is in the fixation plane. Target visibility is smaller than for the No Kernel condition because the four middle elements produce some spatial inhibition that weakens the target boundaries.

The Minimal Kernel condition removes four of the five central elements of the standard mask; leaving only the most central element. This remaining element can produce the binocular false matches needed for shine-through, but produces less spatial competition than the normal 5 Element kernel. As Fig. 20a shows, the model predicts a very weak shine-through effect, with stronger target evidence in the fixation plane. There is enough lateral inhibition to weaken the target boundaries and briefly disinhibit the false matches at the foreground disparity plane. However, without the additional elements of the kernel, the disinhibition is too weak to produce a strong shine-through effect.

In the 4 Lines Gap condition, the top and bottom context elements provide a horizontal boundary at the location of the gap in the standard mask. These horizontal boundaries allow the horizontal boundaries of the standard mask on opposite sides of the gaps to group together. It is interesting to contrast this grouping of horizontal boundaries against the grouping of vertical boundaries in the 4 Lines condition in Fig. 14. The impact of the four lines depends on the overall spatial layout of the rest of the mask. With the grouping of horizontal boundaries and the spatial competition from the 5 Elements of the mask around the target, the 3D LAMINART model predicts that shine-through should occur. Fig. 20a shows that the strongest integrated target evidence for this condition is in the foreground disparity plane.

In the anti-context condition, the standard mask contains only 5 Elements (the kernel), while the top and bottom context elements cover the remaining 20 elements of a 25 Elements mask (anti-context). As in the 4 Lines Gap condition, the horizontal boundaries of the top and bottom context elements group with the horizontal boundaries of the kernel. This grouping promotes shine-through. Fig. 20a shows that the integrated target evidence for this condition is strongest in the foreground disparity plane. It is unlikely that the WCTM would predict shine-through for this condition because the outer elements of the kernel should produce strong inhibition toward the target.

10. Conclusions

The shine-through effect is an important phenomenon because it challenges both spatial and temporal theories of visual perception. For most temporal theories, more mask energy corresponds to stronger masking (see the review in Francis & Herzog, 2004). When shine-through occurs, the expected inhibition of the target instead leads to a changed target percept that occludes the mask. As Herzog and colleagues have pointed out, the existence of the shine-through effect is strong evidence that theories of masking must address the spatial properties of the target and mask.

In addition, Herzog and colleagues have used the shine-through effect to investigate the properties of perceptual grouping in both normal (Herzog & Fahle, 2002) and patient populations (Herzog, Kopmann, & Brand, 2004). Because the shine-through effect is sensitive to the way elements in the visual scene are grouped together, its presence or absence can be used as a quantitative measurement of different perceptual organizations of elements in a scene. The shine-through results thus provide a rich data set for challenging spatial models of perceptual grouping.

The current simulations demonstrate that the 3D LAMINART model explains the properties of shine-through by addressing both spatial and temporal properties of visual computations. The model explains why the mask leads to a perceptual experience of the target occluding the mask and explains why increasing the mask strength, by increasing the number of bars, tends to increase the shine-through effect. In addition, the model explains why the shine-through effect is so sensitive to the perceptual grouping of boundaries. The 3D LAMINART model suggests that the shine-through effect is a natural result of mechanisms involved in depth perception. Indeed, the current simulations have not added any new characteristics to the 3D LAMINART model. All the mechanisms in the model have previously been introduced to explain other aspects of visual perception.

To summarize, the 3D LAMINART model claims that the shine-through effect occurs when the target and mask produce false binocular matches that are consistent with a vernier at the foreground disparity plane. For those false matches to introduce a lasting percept, the mask elements' boundaries must strongly inhibit the target's boundaries at the fixation disparity plane, and the mask element's horizontal boundaries must group together in a way that prevents the target and mask vertical boundaries from grouping.

There are several different ways for the conditions for shine-through to be satisfied (and many more ways for the conditions to not be satisfied). The simulations of previous experimental results demonstrate that it is the boundary grouping properties of these displays that are critical for producing the shine-through effect. In addition, the model makes predictions for novel mask conditions that can be used to test the model's explanation.

The properties of shine-through provide additional evidence for the mechanisms of the 3D LAMINART model. One scientific challenge for working with the 3D LAMINART model has been trying to identify experimental methods that were sensitive enough to measure the subtle effects of the model's hypothesized boundary grouping processes. By connecting the model behavior to the shine-through effect, we anticipate that it will now be possible to generate specific hypotheses that can be tested by the shine-through effect. With such studies, it should now be possible to challenge and refine the details of the model's systems and parameters. Thus, we see the current results not as a general solution to

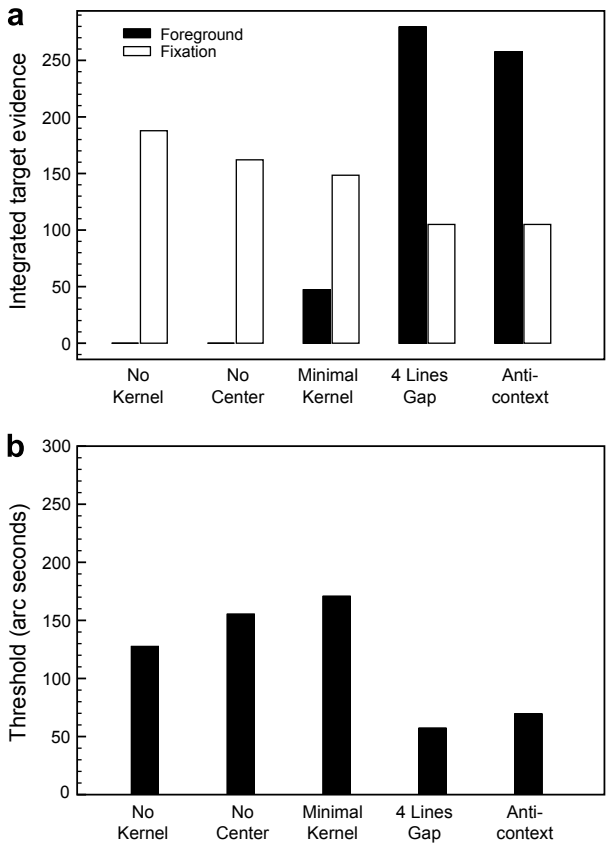

Fig. 20. Predicted target vernier visibility for the masks in Fig. 19. (a) Integrated target evidence across the foreground and fixation planes of the 3D LAMINART model. Large values of integrated target evidence in the foreground disparity plane indicates the presence of shine-through. (b) The largest integrated target evidence across disparity planes can be converted in to a predicted vernier threshold.

the peculiarities of the shine-through effect, but as the start of a rich interaction between data and modeling where each side challenges the other to consider new possibilities.

Appendix A

All of the simulations were programmed in C++ and run on an iMac with a 2.8 GHz processor and 2 GB RAM. With this hardware it took approximately 36 h to go through all of the simulations reported in this paper.

Almost every cell activity is described as a differential equation, which is integrated through time. There are a few exceptions, as described below.

A.1. Stimuli

Each bar element was 20 pixels high and 1 pixel wide. Each pixel separation corresponds to a physical separation of approximately 30 arc sec. The vertical gap between top and bottom elements of the target and standard mask elements was 5 pixels. Mask elements were separated by a gap of 5 pixels in the horizontal direction. The top target vernier element was shifted two pixels to the left, and the bottom target vernier element was shifted two pixels to the right. The top and bottom context elements in Figs. 14 and 19 were 15 pixels high. They were separated from the other mask elements by a gap of 6 pixels. All stimulus elements were bright (arbitrary units of 4) on a dark background (0 units).

The target was always presented for 20 ms (0.4 simulation time units). The mask was always presented for 300 ms (6 simulation time units). The mask immediately followed the offset of the target.

Each pixel in the model is represented by coordinates (i,j). For the input at each pixel, we use superscripts L and R to indicate whether the input is to the Left or Right eye. For all of the stimuli related to shine-through the image pattern to the left and right eye had zero disparity, so $I^L_{ij} = I^R_{ij}$.

In the following sections, we describe the equations for cell activity in each stage of the model. Following the terminology of Cao and Grossberg (2005), we refer to the cells with terms that indicate their neurophysiological counterparts. As much as possible, we describe the equations with the terms and parameters used by Cao and Grossberg (2005).

A.2. Lateral geniculate nucleus

The LGN cells receive on-center, off-surround filtered inputs. The activity of a cell at pixel (i,j) for the left monocular pathway is described by the following differential equation:

$$\frac{dx^L_{ij}}{dt} = \left\{ -\alpha x^L_{ij} + \left(\beta - x^L_{ij}\right) I^L_{ij} - x^L_{ij} \sum_{p,q \neq i,j} G_{pqij} I^L_{pq} \right\} \rho. \tag{2}$$

Here, $\alpha = 10$ sets the rate of passive decay with no input, $\beta = 9.9$ determines an upper limit for the excitatory input, and $\rho = 0.1$ establishes an overall rate of change for the equation. The off-center kernel is defined by:

$$G_{pqij} = \exp\left(-\frac{(p-i)^2 + (q-j)^2}{2\sigma^2} \right), \tag{3}$$

where $\sigma = 3.9$ establishes the spread of the kernel. Each kernel was restricted to a grid of eight by eight pixels.

A.3. V1, Layer 4 simple cells

The V1, Layer 4 simple cells are sensitive to a dark–light contrast polarity (indicated by a + superscript) or a light–dark contrast polarity (indicated by a − superscript) pattern of a preferred (vertical or horizontal) orientation from one eye. The activity of a cell at pixel (i,j) with a preferred orientation k, for the left monocular pathway is given by the following differential equation:

$$\frac{ds^{L+}_{ijk}}{dt} = \left\{ -\alpha s^{L+}_{ijk} + \sum_{p,q \neq i,j} K^{+}_{pqk} \left[x^L_{i+p,j+q} \right]^+ \right\} \rho. \tag{4}$$

Here, $\alpha = 1.0$ sets the rate of passive decay with no input, $\rho = 10$ sets the overall rate of change for the equation, and the notation $[x]^+ = \max(x, 0)$ indicates half-wave rectification. Orientation sensitivity is set by the properties of the oriented filter:

$$K^{+}_{pqk} = \phi \sin\left(\frac{2\pi(r-0.5)}{\tau} \right) \exp\left(-\frac{1}{2} \left[\frac{(p-0.5)^2}{\sigma_p^2} + \frac{(q-0.5)^2}{\sigma_q^2} \right] \right), \tag{5}$$

where $\phi = 4.4$, $\tau = 3\pi$, $\sigma_p = \sigma_q = 0.6$ represent the amplitude and dimensions of the kernel. The term r defines the kernel's orientation. If $r = p$, the orientation is horizontal, when $r = q$ the orientation is vertical. The kernel amplitudes were then normalized so that the maximum value equaled 1.0. The kernel for a light–dark polarity simply inverted the sign of the dark–light kernel.

A.4. V1, Layer 3B monocular simple cells

The V1, Layer 3B monocular simple cells receive inputs from the V1, Layer 4 simple cells and also receive a tonic source of excitation

that subsequently drives an end-cutting process for insuring strong responses to the end of thin lines (Grossberg & Mingolla, 1985a) and generates orientation after-responses that control visible persistence (Francis et al., 1994). As part of the end-cutting process, these cells receive spatial competition from nearby positions of the same preferred orientation and polarity. The activity of a cell at pixel (i,j) with a preferred orientation k, with a dark–light polarity sensitivity (+), for the left monocular pathway is given by the following differential equation:

$$\frac{db^{L+}_{ijk}}{dt} = \left\{ -\alpha b^{L+}_{ijk} + J + 2[s^{L+}_{ijk} - \theta]^+ - v \sum_{p,q} [s^{L+}_{ijk} - \theta]^+ \right\} \rho. \quad (6)$$

Here, $\alpha = 1.0$ sets the rate of passive decay with no input, $\rho = 10$ sets the overall rate of change for the equation, $\theta = 0.4$ is a threshold, $J = 1$ indicates tonic excitatory input, and $v = 0.5$ weights the inhibitory input from nearby neighbors. The spatial competition summation is orientation sensitive. For a horizontal filter it has a width of three pixels and a height of one pixel. The excitatory input from the V1, Layer 4 simple cells is weighted by the number 2 to make it roughly equal in strength to the V1, Layer 3B binocular simple cells (described below).

A.5. V1, Layer 3B monocular simple cell neurotransmitter gates

The V1, Layer 3B monocular simple cells send signals to other cells via a depletable transmitter gate. This, in combination with other types of competition described below, produces a gated dipole circuit (Grossberg, 1972, 1980) that produces orientation after-responses. The available neurotransmitter at pixel (i,j) with a preferred orientation k, with a dark–light polarity sensitivity (+), for the left monocular pathway is given by the following differential equation:

$$\frac{dh^{L+}_{ijk}}{dt} = \left\{ A - (B + b^{L+}_{ijk}) h^{L+}_{ijk} \right\} \rho. \quad (7)$$

Here, $A = 2$ indicates the creation of neurotransmitter, $B = 1$ sets the rate of passive decay of neurotransmitter, $\rho = 0.01$ sets the overall rate of the equation (which is much slower than for cell activities). Signals from the V1, Layer 3B monocular simple cell deplete neurotransmitter stores. At the start of each simulation, every gate is initialized to $h(0) = A/(B + J)$, which is the equilibrium value if only the tonic input to the V1, Layer 3B monocular simple cells is available. The neurotransmitter gates were not part of the simulations used by Cao and Grossberg (2005), who did not investigate the model's dynamic behavior.

A.6. V1, Layer 2/3 monocular complex cells

The V1, Layer 2/3 monocular complex cells receive a variety of excitatory and inhibitory inputs. These cells pool together V1, Layer 3B cells of the same position and orientation preference but opposite polarities. This type of cell also receives inhibitory inputs from cells of the same position but orthogonal orientations, and receives competition from neighboring cells. Define the net input for a cell at pixel (i,j) with a preferred orientation k, for the left monocular pathway as:

$$E^L_{ijk} = \left[(b^{L+}_{ijk} h^{L+}_{ijk} + b^{L-}_{ijk} h^{L-}_{ijk}) - (b^{L+}_{ijK} h^{L+}_{ijK} + b^{L-}_{ijK} h^{L-}_{ijK}) \right]^+, \quad (8)$$

which adds together the gated activity from the V1, Layer 3B cells of the same orientation and different polarities and then subtracts similar signals for the opposite orientation (indexed by K). This subtraction of inputs for the orthogonal orientation produces a gated dipole circuit that generates rebounds of activity at stimulus offset. The net input is rectified, so that only one orientation receives positive input at any moment in time.

The activity of a cell at pixel (i,j) with a preferred orientation k, for the left monocular pathway is given by the following differential equation:

$$\frac{dc^L_{ijk}}{dt} = -\alpha c^L_{ijk} + (\beta - c^L_{ijk})(E^L_{ijk} + \gamma_3 [c^L_{ijk} - \beta_c]^+) - (1 + c^L_{ijk}) \left(\gamma_4 [c^L_{ijK} - \beta_c]^+ + \gamma_5 \sum_{pq} W_{pqijk} [c^L_{pqk} - \beta_c]^+ \right). \quad (9)$$

Here, $\alpha = 20$ sets the rate of passive decay with no input and $\beta = 8$ sets an upper limit for cell activity. The second excitatory term is self-excitation, with $\beta_c = 0.03$ acting as a threshold for such self-excitation and $\gamma_3 = 0.5$ weighting the self-excitation. The terms on the far right correspond to two sources of inhibitory competition. The first describes a competition with a cell of the opposite orientation. Usually this competition is weak since only one orientation receives excitatory input at a time, however, lingering responses can still compete over time. It is weighted by the parameter $\gamma_4 = 5$. The summation term describes competition from spatial neighbors that have the same preferred orientation, which is weighted by $\gamma_5 = 1.5$. The competition weights are described by an elongated Gaussian:

$$W_{pqijk} = \exp\left(-\left[\frac{(p-i)^2}{\sigma_p^2} + \frac{(q-j)^2}{\sigma_q^2} \right] \right). \quad (10)$$

For a horizontal orientation, $\sigma_p = 0.3$ and $\sigma_q = 8$. These were reversed for a vertical orientation. The range of competition was seven by seven pixels.

A.7. V1, Layer 3B interneuron cells

A V1, Layer 3B interneuron cell receives excitatory input from the V1, Layer 4 simple cell with an appropriate disparity shift from one eye, but it also receives inhibitory input from other V1, Layer 3B interneuron cells that code the same position and disparity but different eye input and/or polarity. The activity of a cell at pixel (i,j) with a preferred orientation k, with a dark–light polarity sensitivity (+), that receives excitation from the left eye (L), with a disparity sensitivity of d, is given by the following differential equation:

$$\frac{dq^{L+}_{ijkd}}{dt} = -\gamma_2 q^{L+}_{ijkd} + [s^{L+}_{i+s,jk} - \theta]^+ - \beta \left[q^{L-}_{ijkd} + q^{R+}_{ijkd} + q^{R-}_{ijkd} \right]. \quad (11)$$

Here, $\gamma_2 = 4.5$ sets the rate of passive decay with no input, $\theta = 0.4$ is a threshold for the inputs, $\beta = 4$ weights inhibitory input from all other V1, Layer 3B interneuron cells at the same position, orientation, and disparity. The subscript term $s = 0$ for the fixation disparity plane, indicating that these cells look for matching edges at the same relative positions. For the foreground disparity plane, $s = -1$, which indicates that these cells are part of a circuit that looks for matching edges that are offset by a total of two pixels (one in each eye, relative to the position of the binocular cell). A similar equation for V1, Layer 3B interneurons cells that receive input from the right eye would obey:

$$\frac{dq^{R+}_{ijkd}}{dt} = -\gamma_2 q^{R+}_{ijkd} + [s^{R+}_{i-s,jk} - \theta]^+ - \beta \left[q^{L-}_{ijkd} + q^{L+}_{ijkd} + q^{R-}_{ijkd} \right], \quad (12)$$

which changes the sign of the allelotropic shift s.

A.8. V1, Layer 3B binocular simple cells

A V1, Layer 3B binocular simple cell combines inputs from the V1, Layer 4 simple cells across the two eyes at an appropriate disparity shift, which is represented as s pixels. These cells maintain their sensitivity to a particular contrast polarity, and are part of a competitive circuit that insures the inputs from the two eyes are of roughly the same strength. The activity of a cell at pixel (i,j) with a preferred orientation k, with a dark–light polarity sensitivity (+), and with a disparity sensitivity of d is given by the following differential equation:

$$\frac{db^{B+}_{ijkd}}{dt} = -\gamma_1 b^{B+}_{ijkd} + (1 - b^{B+}_{ijkd})\left([s^{L+}_{i+s,jk} - \theta]^+ + [s^{R+}_{i-s,jk} - \theta]^+\right) - \alpha\left[q^{L+}_{ijkd} + q^{L-}_{ijkd} + q^{R+}_{ijkd} + q^{R-}_{ijkd}\right]. \quad (13)$$

Here, $\gamma_1 = 0.1$ sets the rate of passive decay with no input, $\theta = 0.4$ is a threshold for the inputs, $\alpha = 7.2$ weights inhibitory input from V1, Layer 3B interneuron cells at the same position, orientation, and disparity. The subscript term s equals zero for the fixation disparity plane, indicating that these cells look for matching edges at the same relative positions. For the foreground disparity plane, $s = -1$, which indicates that these cells look for matching edges that are offset by a total of two pixels (one in each eye, relative to the position of the binocular cell).

A.9. V1, Layer 2/3 binocular complex cells

The V1, Layer 2/3 binocular complex cells are similar to the V1, Layer 2/3 monocular cells. Each of these cells pool together V1, Layer 3B binocular simple cells of the same disparity, position, and orientation preference but opposite polarities. It also receives inhibitory inputs from cells of the same position but orthogonal orientations, and receives competition from neighboring cells. Finally, it also receives excitatory feedback from a short-range grouping mechanism. Define the net input for a cell at pixel (i,j) with a preferred orientation k, for the left monocular pathway as:

$$E_{ijkd} = \mu\left[\left([b^{+}_{ijkd} - \theta]^+ + [b^{-}_{ijkd} - \theta]^+\right) - \left([b^{+}_{ijKd} - \theta]^+ + [b^{-}_{ijKd} - \theta]^+\right)\right]^+, \quad (14)$$

which adds together activity from the V1, Layer 3B binocular simple cells of the same orientation, disparity, and different polarities, and then subtracts similar signals for the opposite orientation (indexed by K). Each input is thresholded by the term $\theta = 0.1$. The net input is rectified, so that only one orientation receives positive input at any moment in time, and is weighted by $\mu = 20$.

The activity of a V1, Layer 2/3 binocular complex cell at pixel (i,j) with a preferred orientation k, and disparity d, is given by the following differential equation:

$$\frac{dc_{ijkd}}{dt} = -\alpha c_{ijkd} + (\beta - c_{ijkd})\left(E_{ijkd}(\gamma_1 + \gamma_2 H_{ijkd}) + \gamma_3[c_{ijkd} - \beta_c]^+\right) - (1 + c_{ijkd})\left(\gamma_4[c_{ijKd} - \beta_c]^+ + \gamma_5 \sum_{pq} W_{pqijk}[c_{pqkd} - \beta_c]^+\right). \quad (15)$$

Here, $\alpha = 20$ sets the rate of passive decay with no input and $\beta = 7$ sets an upper limit for cell activity. The parameters $\gamma_1 = \gamma_2 = 1$ weight direct bottom-up input and feedback input from long-range cooperation, as described below. Self-excitation is weighted by $\gamma_3 = 0.5$ and thresholded with $\beta_c = 0.03$. The terms on the far right correspond to two sources of inhibitory competition, as for the V1, Layer 2/3 monocular cells. Here $\gamma_4 = 5$ and $\gamma_5 = 1$.

The term H_{ijkd} describes input from connections within V1, Layer 2/3. The connections define a receptive field that is stretched parallel to the preferred orientation of the cell. It has two lobes on opposite sides of the cell's location. For a horizontal cell the lobes are defined by:

$$W^c_{pqijk1} = \begin{cases} \exp\left(-\left[\frac{(p-i)^2}{\sigma_p^2} + \frac{(q-j)^2}{\sigma_q^2}\right]\right) & \text{if } p < i \\ 0 & \text{otherwise} \end{cases} \quad (16)$$

for the left lobe and

$$W^c_{pqijk2} = \begin{cases} \exp\left(-\left[\frac{(p-i)^2}{\sigma_p^2} + \frac{(q-j)^2}{\sigma_q^2}\right]\right) & \text{if } p > i \\ 0 & \text{otherwise} \end{cases} \quad (17)$$

for the right lobe. Each kernel lobe was subsequently normalized so that its maximum value had a magnitude of one. Together the two lobes covered seven by one pixels, with $\sigma_p = 14$ and $\sigma_q = 0.3$. The σ parameters were swapped for vertically aligned lobes.

The input filtered through the two lobes is defined as:

$$H_{ijkd1} = \sum_{pq} W^c_{pqijk1}[c_{pqkd} - \zeta_c]^+ \quad (18)$$

for the left side, and

$$H_{ijkd2} = \sum_{pq} W^c_{pqijk2}[c_{pqkd} - \zeta_c]^+ \quad (19)$$

for the right side. Here, $\zeta_c = 0$ insures that only positive input values are used.

The activity coming through each lobe is sent to several different cells in order to create a bipole property for the long-range feedback. The bipole property is that positive feedback will only appear when both lobes receive sufficiently strong input. If only one of the lobes receives strong input, there will be no net excitatory feedback. This property is accomplished through two inhibitory interneurons that each receive excitatory input from one of the lobes. These interneurons inhibit each other, so that if both lobes are active then neither inhibitory interneurons is very strong. Such weakening prevents inhibition at the V1, Layer 2/3 binocular cell and allows input from the lobes to strengthen the V1, Layer 2/3 binocular cell response. The activity of the interneuron receiving input from the first lobe is:

$$\frac{ds^c_{ijkd1}}{dt} = -s^c_{ijkd1} + H_{ijkd1} - \eta s^c_{ijkd1}[s^c_{ijkd2}]^+, \quad (20)$$

where $\eta = 1$ weights shunting inhibition from any positive activity in the other interneuron. A similar equation exists for the other interneuron with all references to lobe 1 replaced by lobe 2, and vice-versa. This pair of equations is computationally difficult to integrate through time. Moreover, according to Cao and Grossberg (2005), these equations operate on a much faster time scale than the other differential equations in the network. As a result, we solved each equation at its steady state. For the first interneuron this is:

$$s^c_{ijkd1} = \frac{-B_1 + \sqrt{B_1^2 + 4\eta H_{ijkd1}}}{2\eta}, \quad (21)$$

where

$$B_1 = 1 + \eta(H_{ijkd2} - H_{ijkd1}). \quad (22)$$

The equation for the second interneuron is similar, with subscript references to lobes 1 and 2 swapped.

The net input to the V1, Layer 2/3 binocular cell is then the sum of input among the lobes and inhibition from each of the interneurons:

$$H_{ijkd} = \left[H_{ijkd1} + H_{ijkd2} - s^c_{ijkd1} - s^c_{ijkd1}\right]^+. \quad (23)$$

Any negative values are rectified to zero.

A.10. V2, Layer 4 cells

The V2, Layer 4 cells combine binocular input from the V1, Layer 2/3 binocular cells with appropriate inputs from the V1, Layer 2/3 monocular cells. The monocular cells project to V2, Layer 4 cells that are at pixel positions at each monocular cell's line-of-sight. That is, a monocular cell projects to a binocular cell at a position that corresponds to where such monocular edges would contribute to a disparity match.

The activity of a V2, Layer 4 cell at pixel (i,j) with a preferred orientation k, and disparity d, is given by the following differential equation:

$$\frac{dv_{ijkd}}{dt} = \left\{-v_{ijkd} + \alpha h(c_{ijkd} - \theta) + \beta h\left([c^L_{i+s,jk} - \theta]^+ + [c^R_{i-s,jk} - \theta]^+\right)\right\}\rho, \tag{24}$$

where $\alpha = 2.6$ weights the binocular input, $\beta = 0.3$ weights the monocular input, $\theta = 0.06$ is a threshold for the inputs, $\rho = 2$ sets the rate of the equation, and the function $h(x)$ is a hard threshold:

$$h(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{25}$$

In the simulations of Cao and Grossberg (2005), these V2, Layer 4 cells also received excitatory and inhibitory feedback from V2 monocular surfaces. This feedback proved computationally difficult to simulate dynamically, and it was not included in the current simulations. Fortunately, it does not appear to be critical for explaining shine-through. For the present simulations, we also introduced a second intermediary V2, Layer 4 interneuron that introduced orientational and spatial competition. This competition was not part of the simulations used by Cao and Grossberg (2005) but has been part of precursors to the 3D LAMINART model for explaining both spatial processing (Grossberg & Mingolla, 1985a, 1985b) and backward masking (Francis, 1997). The activity of a V2, Layer 4 interneuron at pixel (i,j) with a preferred orientation k, and disparity d, is given by the following differential equation:

$$\frac{dy_{ijkd}}{dt} = -y_{ijkd} + \beta\left[[v_{ijkd}]^+ - [v_{ijKd}]^+\right]^+ - y_{ijkd} v \sum_{pq} L_{pqijk} v_{pqkd}, \tag{26}$$

where K is the orthogonal orientation of k. The parameter $\beta = 5$ scales the net excitatory input and $v = 0.9$ weights a spatial competition from neighboring cells of the same orientation. The inhibitory kernel L_{pqijk} is defined as a modified elongated Gaussian:

$$L_{pqijk} = \begin{cases} \exp\left(-\left[\frac{(p-i)^2}{\sigma_p^2} + \frac{(q-j)^2}{\sigma_q^2}\right]\right) & \text{if } q \neq j, \\ 0 & \text{if } q = j. \end{cases} \tag{27}$$

For a horizontal orientation, $\sigma_p = 2$ and $\sigma_q = 6$. These were reversed for a vertical orientation. All values were subsequently normalized so that the maximum value was equal to one. The range of competition was five by thirteen pixels, with the longer side in the orthogonal orientation of the cell's orientation preference.

A.11. V2, Layer 2/3 interneurons

The V2, Layer 2/3 interneurons receive excitatory and inhibitory inputs from V2, Layer 4 cells. These cells also receive excitatory grouping signals, similar to the grouping signals for the V1, Layer 2/3 cells, that combine inputs from disparate V2, Layer 4 interneurons. These grouping signals are modulated by inhibitory signals from V2, Layer 2/3 cells (described below) of the orthogonal orientation.

Define the direct input for a cell at pixel (i,j) with a preferred orientation k, as:

$$E_{ijkd} = \gamma_1\left[([y_{ijkd}]^+ - [y_{ijKd}]^+\right]^+, \tag{28}$$

which contrasts activity from the V2, Layer 4 interneurons of the opposite orientations and the same position and disparity. The net input is rectified, so that only one orientation receives positive input at any moment in time, and is weighted by $\gamma_1 = 1.3$.

The activity of a V2, Layer 2/3 interneuron at pixel (i,j) with a preferred orientation k, and disparity d, is given by the following differential equation:

$$\frac{dg_{ijkd}}{dt} = \left\{-\alpha g_{ijkd} + (\beta - g_{ijkd})E_{ijkd}\left[\gamma_2 H_{ijkd} - \gamma_6 N_{ijkd}\right]^+\right\}\rho. \tag{29}$$

Here, $\rho = 8$ influences the overall rate of the equation, $\alpha = 30$ sets the rate of passive decay with no input, and $\beta = 10$ sets an upper limit for cell activity. The parameter $\gamma_2 = 8$ weights excitatory grouping signals and $\gamma_6 = 100$ weights inhibition from the orthogonal orientation that prevents grouping.

As for the V1, Layer 2/3 cells, the term H_{ijkd} describes input from other interneurons within V1, Layer 2/3. The connections define a receptive field that is stretched parallel to the preferred orientation of the cell. It has two lobes on opposite sides of the cell's location. For a horizontal cell the lobes are defined by:

$$W^c_{pqijk1} = \begin{cases} \exp\left(-\left[\frac{(p-i)^2}{\sigma_p^2} + \frac{(q-j)^2}{\sigma_q^2}\right]\right) & \text{if } p < i \\ 0 & \text{otherwise} \end{cases} \tag{30}$$

for the left lobe and

$$W^c_{pqijk2} = \begin{cases} \exp\left(-\left[\frac{(p-i)^2}{\sigma_p^2} + \frac{(q-j)^2}{\sigma_q^2}\right]\right) & \text{if } p > i \\ 0 & \text{otherwise} \end{cases} \tag{31}$$

for the right lobe. Each kernel lobe was subsequently normalized so that its maximum value had a magnitude of one. Together the two lobes covered eleven by one pixels, with $\sigma_p = 5.5$ and $\sigma_q = 0.3$. The σ parameters were swapped for vertically aligned lobes.

The input filtered through the two lobes is defined as:

$$H_{ijkd1} = \sum_{pq} W^c_{pqijk1}[g_{pqkd} - \zeta_J]^+ \tag{32}$$

for the left side, and

$$H_{ijkd2} = \sum_{pq} W^c_{pqijk2}[g_{pqkd} - \zeta_J]^+ \tag{33}$$

for the right side. Here, $\zeta_J = 0.01$.

Just as for the V1, Layer 2/3 cells, the activity of the lobes project to interneurons that create a bipole property for the V2, Layer 2/3 interneurons. These interneurons are defined in the same way as for the V1, Layer 2/3 cells. The net input to the V2, Layer 2/3 interneuron is then the sum of input among the lobes and the inhibition from each of the interneurons:

$$H_{ijkd} = \left[H_{ijkd1} + H_{ijkd2} - s^c_{ijkd1} - s^c_{ijkd1}\right]^+. \tag{34}$$

Any negative values are rectified to zero.

Any grouping excitation can be inhibited by the presence of activity from the orthogonal orientation among V2, Layer 2/3 cells (described below). The inhibitory kernel samples activities parallel to the preferred orientation of the V2, Layer 2/3 interneurons. The range of sampling is larger than the excitatory kernels, and for the present simulations it spans the entire range of the input image (a row for a horizontally tuned cell and a column for a vertically tuned cell). Thus for a horizontally tuned V2, layer 2/3 interneuron,

$$N_{ijkd} = \sum_{p} u_{pjKd}, \tag{35}$$

where the summation is across all cells in row j and K indicates the orientation orthogonal to k.

A.12. V2, Layer 2/3 cells

The V2, Layer 2/3 cells receive excitatory inputs from V2, Layer 4 interneurons and V2, Layer 2/3 interneurons of the same disparity, inhibitory inputs from V2, Layer 2/3 cells at other disparities, and spatial competition from V2, Layer 4 interneurons. The activity of a V2, Layer 2/3 cell at pixel (i,j) with a preferred orientation k, and disparity d, is given by the following differential equation:

$$\frac{du_{ijkd}}{dt} = \left\{ -\alpha u_{ijkd} + \beta(\gamma_1 E_{ijkd} + \gamma_2 y_{ijkd}) - G_{ijkd} - u_{ijkd} v \sum_{pq} L_{pqijk} y_{pqkd}, \right\} \rho. \tag{36}$$

Here $\alpha = 1$ sets the rate of passive decay, $\rho = 10$ sets the overall rate of the equation, $\beta = 0.1$ weights the net excitatory input, $\gamma_1 = 1.5$ scales pooled input from nearby V2, Layer 4 interneurons, $\gamma_2 = 8$ weights direct excitatory input from a V2, Layer 4 interneuron, G_{ijkd} pools inhibition from V2, Layer 2/3 cells at other disparity planes, $v = 0.35$ weights spatial competition from nearby V2, Layer 4 interneurons at the same orientation and disparity, the kernel L_{pqijk} is the same as defined in Eq. (27).

E_{ijkd} pools signals from V2, Layer 4 interneurons across regions defined by the V2, Layer 2/3 interneurons. For a horizontally tuned V2, Layer 2/3 cell, we pool together the above threshold activities of V2, Layer 4 interneurons that have a direct path among V2, Layer 2/3 interneurons to the cell at position (i,j). Algorithmically, for a horizontally tuned cell this pooling is accomplished by summing connected activities on the left and right side of pixel (i,j). To the left of position (i,j), we find the first pixel that has V2, Layer 2/3 interneuron activity below a threshold value:

$$L = \arg \max_{p, p-i \leqslant 0} [g_{pjkd} - \theta]^+, \tag{37}$$

where $\theta = 0.01$. L is the x coordinate of the closest pixel to the left of pixel (i,j) that has a V2, Layer 2/3 interneuron activity that is not above threshold. Likewise, define for the right side:

$$R = \arg \min_{p, p-i \geqslant 0} [g_{pjkd} - \theta]^+, \tag{38}$$

which gives the x coordinate of the closest pixel to the right of pixel (i,j) that has a V2, Layer 2/3 interneuron activity that is not above threshold.

The pooling is then across V2, Layer 4 interneurons at positions between L and R, but not including pixel (i,j). First count up the number of V2, Layer 4 interneurons that contribute non-zero input:

$$C = \sum_{p=i-1}^{L+1} h(v_{pjkd}) + \sum_{p=i+1}^{R-1} h(v_{pjkd}), \tag{39}$$

where $h(x)$ is defined in Eq. (25). Net input to a V2, Layer 2/3 cell is non-zero only if enough ($\beta_b = 12$) V2, Layer 4 interneurons are being pooled.

$$E_{ijkd} = \begin{cases} \sum_{p=i-1}^{L+1} v_{pjkd} + \sum_{p=i+1}^{R-1} v_{pjkd} & \text{if } C > \beta_b, \\ 0 & \text{otherwise.} \end{cases} \tag{40}$$

This excitatory pooling is offset by line-of-sight inhibition from V2, Layer 2/3 cells at other disparity planes. Thus, a horizontally tuned cell at pixel (i,j) will pool other V2, Layer 2/3 cells of the orthogonal orientation across row j:

$$G_{ijkd} = \gamma_3 \sum_{d'} M_{d'd} [u_{(i+s'-s)jkd'} - \beta_g]^+, \tag{41}$$

where $\beta_g = 0.2$ is a threshold and $M_{d'd}$ defines the strength of inhibition from disparity plane d' to disparity plane d. These disparity planes have allelotropic shifts s' and s, respectively, which together identify the pixels in disparity plane d' that are at the same line-of-sight position as pixel (i,j) in disparity plane d. The competition between disparity planes is biased toward the fixation plane, so only the fixation disparity plane inhibits the foreground disparity plane:

$$M_{d'd} = \begin{cases} 1.1 & \text{for } d' \text{ as the fixation plane and } d \text{ as the foreground,} \\ 0 & \text{otherwise.} \end{cases} \tag{42}$$

The strength of this inhibition varied for vertical orientations ($\gamma_3 = 2.75$) and horizontal orientations ($\gamma_3 = 0.05$). To better demonstrate the model interactions responsible for removing false binocular matches with static stimuli, the simulations for Figs. 3 and 4 used $M_{d'd} = 2$ and $M_{dd'} = 0.55$.

A.13. V2, Thin stripe monocular filling-in cells

There are multiple V2, Thin stripe disparity planes that all receive brightness signals from LGN cells of a single eye and receive binocular boundary signals from the V2, Layer 2/3 cells of a matching disparity. The activity of a V2, Layer 2/3 cell at pixel (i,j) and disparity d for the left eye, is given by the following differential equation:

$$\frac{dF^L_{ijd}}{dt} = \left\{ -\alpha F^L_{ijd} + \sum_{p,q \in N_{ij}} (F^L_{pqd} - F^L_{ijd}) \Phi_{pqijd} + \gamma_1 X^L_{ijd} \right\} \rho, \tag{43}$$

where $\alpha = 10$ sets the rate of passive decay, $\gamma_1 = 1000$ weights the input from LGN cells, and $\rho = 10$ influences the overall rate of the equation.

The set N_{ij} defines a set of nearest neighbor pixels:

$$N_{ij} = \{(i-1,j), (i+1,j), (i,j-1), (i,j+1)\}. \tag{44}$$

The input to this cell, X^L_{ijd} is derived from the LGN cells, with appropriate shifts for the receiving cell's disparity.

$$X^L_{ijd} = \begin{cases} [x^L_{(i+s)j} - \eta]^+ & \text{if } [x^L_{(i+s)j} - \eta]^+ > 0 \\ v & \text{otherwise.} \end{cases} \tag{45}$$

Here $\eta = 0.0001$ is a threshold and the value $v = -10$ acts to remove brightness signals that are not contained within a set of active and connected V2, Layer 2/3 cells.

The Φ_{pqijd} coefficients are modified by the V2, Layer 2/3 cells, such that strong activity at a cell makes the coefficient small. This blocks the spread of brightness information among the neighboring filling-in cells. The blocking is orientation specific, so that vertically tuned V2, Layer 2/3 cells ($k = 1$) block brightness information from spreading horizontally, but allow brightness information to spread vertically. Likewise, horizontally tuned V2, Layer 2/3 cells ($k = 0$) block brightness information from spreading vertically, but allow brightness information to spread horizontally. The coefficients that allow brightness information to spread from pixel (i,j) one pixel to the left $(p,q) = (i-1,j)$ is defined as:

$$\Phi_{pqijd} = \begin{cases} 0 & \text{if } [u_{i(j+1)1d} - \beta_g]^+ + [u_{ij1d} - \beta_g]^+ > 0 \\ \delta & \text{otherwise,} \end{cases} \tag{46}$$

where $\beta_g = 0.2$ is the same threshold as in Eq. (41) for output of the V2, Layer 2/3 cells, and $\delta = 200,000$ indicates a very strong flow of brightness information between cells that do not have boundaries between them. The connection between pixel (i,j) with the cell to the right, at position $(p,q) = (i+1,j)$, has a coefficient defined as:

$$\Phi_{pqijd} = \begin{cases} 0 & \text{if } [u_{(i+1)j1d} - \beta_g]^+ + [u_{(i+1)(j+1)1d} - \beta_g]^+ > 0 \\ \delta & \text{otherwise.} \end{cases} \tag{47}$$

The connection between pixel (i,j) with the cell above, at position $(p,q) = (i,j-1)$, has a coefficient defined as:

$$\Phi_{pqijd} = \begin{cases} 0 & \text{if } [u_{ij0d} - \beta_g]^+ + [u_{(i+1)j0d} - \beta_g]^+ > 0, \\ \delta & \text{otherwise.} \end{cases} \quad (48)$$

where the index for orientation has changed to $k = 0$ to indicate that horizontal boundaries block brightness information from flowing vertically.

The connection between pixel (i,j) with the cell below, at position (p,q)=(i,j+1), has a coefficient defined as:

$$\Phi_{pqijd} = \begin{cases} 0 & \text{if } [u_{i(j+1)0d} - \beta_g]^+ + [u_{(i+1)(j+1)0d} - \beta_g]^+ > 0, \\ \delta & \text{otherwise.} \end{cases} \quad (49)$$

It is quite difficult to integrate Eq. (43) through time because it requires a much finer time scale than the other equations. In many simulations (e.g., Cao & Grossberg, 2005), a equilibrium solution is found by iterating the steady-state solutions to the equations:

$$F^L_{ijd}(\infty) = \frac{\gamma_1 X^L_{ijd} + \sum_{p,q\in N_{ij}} F^L_{pqd}(\infty)\Phi_{pqijd}}{\alpha + \sum_{p,q\in N_{ij}} \Phi_{pqijd}}. \quad (50)$$

In the current simulations we are interested in the dynamic values, which the steady-state solution does not capture. As a compromise between the needs of tracking the system dynamics and making the computations tractable, we estimated the equilibrium solution for Eq. (50) by iterating the equations for 2000 steps or until the maximum change in a step was less than 0.0001. The resulting equilibrium value was used to take a step in time with the following equation:

$$F^L_{ijd}(t+\Delta t) = F^L_{ijd}(\infty) + \left(F^L_{ijd}(t) - F^L_{ijd}(\infty)\right)\exp(-\Delta t \alpha \rho), \quad (51)$$

which describes exponential growth or decay from the value at time t toward the new limit value at time $t + \Delta t$. We used $\Delta t = 0.01$.

A.14. V4, Binocular filling-in cells

The V4 cells receive lightness signals from the V2 Thin stripes. The activity of a V4 cell at pixel (i,j) and disparity d is given by the following differential equation:

$$\frac{dZ_{ijd}}{dt} = -\alpha Z_{ijd} + \sum_{p,q\in N_{ij}} (Z_{pqd} - Z_{ijd})\Phi_{pqijd} + \gamma_1 X_{ijd}, \quad (52)$$

where $\alpha = 1$, $\gamma_1 = 15$, and the Φ_{pqijd} coefficients are the same as defined for the V2, Thin stripe cells, except that $\delta = 2,000,000$ was used. The input to a V4 cell is a combination of brightness signals from the V2, Thin stripe cells:

$$X_{ijd} = F^L_{ijd} + F^R_{ijd}. \quad (53)$$

A.15. Template matching cells

Evidence of the target vernier, and of an anti-vernier that was shifted in the opposite direction, were computed by template matching cells. Pixels of the target template, T_{ij} took binary values (zero or one). The target vernier offset was to the left for the top half and to the right for the bottom half. The top half of the target template had pixels with the value one that covered the space between the center of the display and 6 pixels to the left (the spacing between elements of the mask grating). The bottom half of the target template took the value one for the space between the center of the display and 6 pixels to the right. The anti-vernier template was symmetric to the vernier template around the central vertical axis.

The evidence for a given template, m (0 for the target vernier and 1 for the anti-vernier), across disparity plane d was calculated as

$$E_{md} = \sum_{ij} T_{ij}[Z_{ijd} - \theta]^+, \quad (54)$$

where $\theta = 0.05$ is a threshold for binocular filled-in activity. The differences $E_{0d} - E_{1d}$ of evidence for the target vernier and anti-vernier are given in Fig. 6 for the foreground and fixation disparity planes.

The integral of the evidence difference for a given disparity, d, is calculated as

$$v_d = \int_0^{T_0} (E_{0d} - E_{1d})d\tau, \quad (55)$$

which simply adds up the evidence difference up to time T_0, which is the end of the simulation. This value is compared to the experimental data, or converted to a threshold with Eq. (1). The threshold calculations used $a = 0.013739$ and $b = 138.1528$.

References

Ansorge, U., Breitmeyer, B. G., & Becker, S. I. (2007). Comparing sensitivity across different processing measures under metacontrast masking conditions. *Vision Research, 47*, 3335–3349.

Baloch, A. A., & Grossberg, S. (1997). A neural model of high-level motion processing: Line motion and formation dynamics. *Vision Research, 37*, 3037–3059.

Breitmeyer, B. (1984). *Visual masking: An integrative approach*. New York: Oxford University Press.

Breitmeyer, B., & Öğmen, H. (2006). *Visual masking: Time slices through conscious and unconscious vision* (2nd ed.). Oxford University Press.

Cao, Y., & Grossberg, S. (2005). A laminar cortical model of stereopsis and 3-D surface perception: Closure and da Vinci stereopsis. *Spatial Vision, 18*, 515–578.

Craft, E., Schütze, H., Niebur, E., & von der Heydt, R. (2007). A neural model of figure-ground organization. *Journal of Neurophysiology, 97*, 4310–4326.

Di Lollo, V., Enns, J. T., & Rensink, R. A. (2000). Competition for consciousness among visual events: The psychophysics of reentrant visual processes. *Journal of Experimental Psychology: General, 129*, 481–507.

Enns, J. T., & Di Lollo, V. (2000). What's new in visual masking. *Trends in Cognitive Sciences, 4*, 345–352.

Francis, G. (1997). Cortical dynamics of lateral inhibition: Metacontrast masking. *Psychological Review, 104*, 572–594.

Francis, G. (2003). Developing a new quantitative account of backward masking. *Cognitive Psychology, 46*, 198–226.

Francis, G. (2007). What should a quantitative model of masking look like and why would we want it? *Advances in Cognitive Psychology, 3*, 21–31.

Francis, G., & Cho, Y. S. (2008). Effects of temporal integration on the shape of visual backward masking functions. *Journal of Experimental Psychology: Human Perception & Performance, 34*, 1116–1128.

Francis, G., & Grossberg, S. (1996). Cortical dynamics of boundary segmentation and reset: Persistence, afterimages, and residual traces. *Perception, 25*, 543–567.

Francis, G., Grossberg, S., & Mingolla, E. (1994). Cortical dynamics of feature binding and reset: Control of visual persistence. *Vision Research, 34*, 1089–1104.

Francis, G., & Herzog, M. (2004). Testing quantitative models of backward masking. *Psychonomic Bulletin & Review, 11*, 104–112.

Francis, G., & Rothmayer, M. (2003). Interactions of afterimages for orientation and color: Experimental data and model simulations. *Perception & Psychophysics, 65*, 508–522.

Geisler, W. S., & Super, B. J. (2000). Perceptual organization of two-dimensional patterns. *Psychological Review, 107*, 677–708.

Grossberg, S. (1972). A neural theory of punishment and avoidance: II. Quantitative theory. *Mathematical Biosciences, 15*, 253–285.

Grossberg, S. (1980). How does a brain build a cognitive code? *Psychological Review, 87*, 1–51.

Grossberg, S. (1997). Cortical dynamics of three-dimensional figure-ground perception of two-dimensional figures. *Psychological Review, 104*, 618–658.

Grossberg, S., & Kelly, F. J. (1999). Neural dynamics of binocular brightness perception. *Vision Research, 39*, 3796–3816.

Grossberg, S., & Hong, S. (2006). A neural model of surface perception: Lightness, anchoring, and filling-in. *Spatial Vision, 19*, 263–321.

Grossberg, S., & Howe, P. D. L. (2003). A laminar cortical model of stereopsis and three-dimensional surface perception. *Vision Research, 43*, 801–829.

Grossberg, S., Kuhlmann, L., & Mingolla, E. (2007). A neural model of 3D shape-from-texture: Multiple-scale filtering, boundary grouping, and surface filling-in. *Vision Research, 47*, 634–672.

Grossberg, S., & Mingolla, E. (1985a). Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading. *Psychological Review, 92*, 173–211.

Grossberg, S., & Mingolla, E. (1985b). Neural dynamics of perceptual grouping: Textures, boundaries, and emergent segmentations. *Perception & Psychophysics, 38*, 141–171.

Grossberg, S., & Yazdanbakhsh, A. (2005). Laminar cortical dynamics of 3D surface perception: Stratification, transparency, and neon color spreading. *Vision Research, 45*, 1725–1743.

Hermens, F., & Herzog, M. H. (2007). The effects of the global structure of the mask in visual backward masking. *Vision Research, 47*, 1790–1797.

Hermens, F., Herzog, M. H., Luksys, G., Gerstner, W., & Ernst, U. (2008). Modeling spatial and temporal aspects of visual backward masking. *Psychological Review, 115*, 83–100.

Herzog, M. H. (2007). Spatial processing and visual backward masking. *Advances in Cognitive Psychology, 3*, 85–92.

Herzog, M. H., Dependahl, S., Schmonsees, U., & Fahle, M. (2004). Valences in contextual vision. *Vision Research, 44*, 3131–3143.

Herzog, M. H., Ernst, U., Etzold, A., & Eurich, C. (2003). Local interactions in neural networks explain global effects in the masking of visual stimuli. *Neural Computation, 15*, 2091–2113.

Herzog, M. H., & Fahle, M. (2002). Effects of grouping in contextual modulation. *Nature, 415*, 433–436.

Herzog, M. H., Fahle, M., & Koch, C. (2001). Spatial aspects of object formation revealed by a new illusion, shine-through. *Vision Research, 41*, 2325–2335.

Herzog, M. H., & Koch, C. (2001). Seeing properties of an invisible object: Feature inheritance and shine-through. *Proceedings of the National Academy for Sciences of the United States of America, 98*, 4271–4275.

Herzog, M. H., Kopmann, S., & Brand, A. (2004). Intact figure-ground-segmentation in schizophrenia. *Psychiatry Research, 129*, 55–63.

Hochberg, J. E. (1971). Perception. In J. W. Kling & L. A. Riggs (Eds.), *Experimental psychology* (3rd ed., pp. 396–550). New York: Holt, Rinehart and Winston.

Kelly, F. J., & Grossberg, S. (2000). Neural dynamics of 3-D surface perception: Figure-ground separation and lightness perception. *Perception & Psychophysics, 62*, 1596–1619.

Kim, H., & Francis, G. (2000). Perceived motion in complementary afterimages: Verification of a neural network theory. *Spatial Vision, 13*, 67–86.

McKee, S. P., Bravo, M. J., Taylor, D. G., & Legge, G. E. (1994). Stereo matching precedes dichoptic masking. *Vision Research, 34*, 1047–1060.

McLoughlin, N. P., & Grossberg, S. (1998). Cortical computation of stereo disparity. *Vision Research, 38*, 91–99.

Poggio, G. F. (1991). Physiological basis of stereoscopic vision. In *Vision and Visual Dysfunction Binocular vision* (pp. 224–238). Boston, MA: CRC.

Rock, I. (1993). The logic of "the logic of perception". *Italian Journal of Psychology, 20*, 841–867.

Ross, W., Grossberg, S., & Mingolla, E. (2000). Visual cortical mechanisms of perceptual grouping: Interacting layers, networks, columns, and maps. *Neural Networks, 13*, 571–588.

Rubin, E. (1915). *Synoplevde figurer*. Copenhagen: Glydendalske.

Weisstein, N. (1972). Metacontrast. In D. Jameson & L. Hurvich (Eds.), *Handbook of sensory physiology*. *Visual psychophysics* (Vol. 7, pp. 233–272). Berlin: Springer-Verlag. No. 4.

Wilson, H. R., & Cowan, J. D. (1973). A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik, 13*, 55–80.